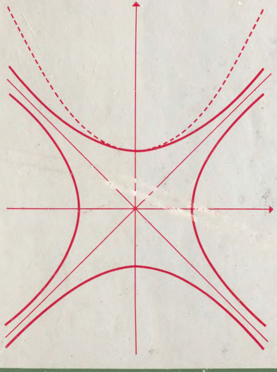
SPECIAL *THEORY* OF *RELATIVITY*

V. A. UGAROV



ИЗДАТЕЛЬСТВО • МИР •

СПЕЦИАЛЬНАЯ ТЕОРИЯ ОТНОСИТЕЛЬНОСТИ

В. А. УГАРОВ

ИЗДАТЕЛЬСТВО «НАУКА»

МОСКВА

SPECIAL *THEORY* OF *RELATIVITY*

V. A. UGAROV

TRANSLATED
FROM
THE RUSSIAN
BY
YURI
ATANOV

MIR PUBLISHERS
MOSCOW

First published 1979
Revised from the 1977 Russian edition

На английском языке

PREFACE

It gives me pleasure to thank B. M. Bolotovsky and S. N. Stolyarov who have written §§ 6.14, 6.15 of this book. I wish to express my special gratitude to V. L. Ginzburg. This book quotes many things which I learned at the seminar led by him. A few questions were discussed with him directly; in particular, the problem of an energy-momentum-tension tensor should be mentioned. Finally, V. L. Ginzburg has written the article "Who Developed the Special Theory of Relativity, and How?" to be published in this book (Supplement I). In my opinion, this article gives very precise answers to questions which would be met by anyone interested in the history of the STR evolution. I feel myself honoured to have this article included in the book.

The author

CONTENTS

CHAPTER I

CLASSICAL MECHANICS AND THE PRINCIPLE OF RELATIVITY

§ 1.1. A coordinate system and a reference frame in classical mechanics. All natural phenomena happen in space and in the course of time, and an element of any phenomenon is something occurring at a given moment of time and at a given point in space. In the special theory of relativity * it is customary to refer to that "something" taking place at a given point and at a given moment of time (in fact, something concentrated in a sufficiently small volume of space and limited by a small time interval) as an event. This definition shows that concrete features of an event may be very different. That is why it is usual to indicate that "the event consists in ...". The examples of events can be the emission of a light signal from a certain point in space at a certain moment of time, or the presence of a moving particle (a material point) at a given point in space and at a given moment of time.

When an event is realized, one says that it "happened" (or is happening, or will happen). Any physical phenomenon represents a sequence of events. A description of a separate event serves as a basis for the description of any phenomenon and therefore we begin with the description of a separate event.

To characterize a point in space where an event occurred, every point in space has to be labelled before specific physical phenomena are analysed. But space is uniform and isotropic and this implies that all points in space and all directions in it are equal. It should be pointed out at once that we deal here with the free space, or vacuum. The investigation of physical phenomena *in vacuo* is of prime importance for the special theory of relativity. Even though vacuum is a complex physical system, it is sufficient for our purpose to assume that in the space domain which we take for vacuum, no substance possessing a finite rest mass is practically present and gravitational and electric fields are not too strong.

But even when all points in space are equal, one can still single out a certain point by placing a material object, i.e. an

* Hereinafter the complete term "special theory of relativity" will be sometimes abbreviated as STR.

object having a finite rest mass, in it. Points in space are usually labelled by means of a coordinate system. With the help of the material object we distinguish a point which is the origin of coordinates The simplest coordinate system is the Cartesian system. Its construction begins with the tracing of three mutually perpendicular straight lines, i.e. the coordinate X, Y, Z axes. In terms of physics, however, these are not just abstract straight lines. Theoretically, the coordinate axes are rigid non-deformable* solids. By the way, instruments, standards and other objects of a given reference frame will be always fixed to them and therefore it should be borne in mind that a physical coordinate system is always a material object.

In the Cartesian coordinate system points are quite easy to label. From any point M in space one can construct the perpendiculars to the X, Y, Z axes or, in other words, project this point on the coordinate axes. Having measured the distances of the point projections from the origin along the X, Y, Z axes by means of the chosen scale, we obtain the numbers x, y, z, which are called the Cartesian coordinates of the point. The distances can be measured via the step-by-step transposition of a unit scale along the axis from the origin to the point projection on the axis. In fact, such a procedure used for length measurements in everyday life can also be used for determining the length of a stretch or an object if it is at rest in a given coordinate system. As we shall see later, the special theory of relativity furnishes a very convenient method of measuring distances without recourse to rigid scales and their step-by-step transposition (see Chapter 2). Both methods are equivalent, of course.

Thus through the introduction of the Cartesian coordinate system every point in space acquires three numbers, that is the three Cartesian coordinates x, y, z. The principal objective of physics, however, is to study motion. Although mechanical motion is the simplest type of motion, its description requires time measurements and therefore the coordinate system has to be of necessity supplemented by a clock. This clock is needed to register the occurrence of events at various points in space. How many clocks are needed?

In classical mechanics they do not usually hesitate over the answer to this question and tacitly assume that one clock resting in a given coordinate system is enough. It is useful to find out what this assumption implies. Let the clock be located at the origin of the coordinate system. Events may happen at any points

* The STR negates the existence of absolute solids (see Chapter 8) but for the coordinate axes it is just sufficient not to be very elastic.

in space including those removed far enough from the origin. Then how can the clock, removed from the place where an event happens, register that event? Obviously, just at the moment the event occurs a certain signal has to be sent from the place of occurrence of the event to the clock located at the origin. If the velocity of the signal is finite, it will reach the clock some time after the onset of the event, and the time lag will depend on the distance between the point where the event occurred and the clock. In classical mechanics, however, it is assumed that basically there may be signals propagating infinitely fast. It is obvious that in this case one clock fixed rigidly to any point in the coordinate system will be enough.

It is implied that the onset of an event is registered as follows: at the moment of an event occurring at any point in space a signal is sent from that point to the clock, and the time of its arrival is thus the time of the onset of the event (the velocity of the signal is infinite!). The assumption concerning the infinitely fast signals applies, of course, not only to the registration of events. In Newtonian mechanics it is incorporated intrinsically: interactions between bodies are transmitted infinitely fast (see § 1.4).

Modern physics, however, claims that all signals (interactions) are transmitted at a finite speed; in other words, there is a finite velocity of interaction transmission. How can this fact be reconciled with the evidence that Newtonian mechanics based on the assumption about the infinitely fast signals copes excellently with many problems (for example, calculates superbly the motion of planets in the solar system)? The answer to this question is very simple. The ultimate speed at which a signal, or an interaction, is transmitted is very great. According to the contemporary ideas it is the speed of electromagnetic waves *in vacuo*, which is equal to approximately $3 \cdot 10^8$ m/s. It follows that as far as velocities of objects to be considered are essentially less than that of light *in vacuo* and characteristic distances are such that the time of light propagation along them is negligibly small, Newtonian mechanics is correct and one clock is enough to register the time of events. Yet it is at once clear that in the case of a fast motion ($v \approx c$) and extended systems a time of an event has to be registered otherwise and the whole science of mechanics has to be based on different premises. In fact, this is just what the special theory of relativity does when it explicitly takes into account the finite velocity of the interaction transmission.

Now let us get back to the classical pattern. A reference frame is formed by a reference object with a coordinate system, a set of length standards and a clock fixed rigidly to a reference object. In physics a reference frame is always implied since any

measurement taken by an instrument produces a result that is related to the reference frame in which this instrument is at rest.

§ 1.2. The choice of a reference frame. To tackle concrete problems, we choose a convenient reference frame and a convenient coordinate system. How does this opportunity of choice come about? As to a clock, in classical mechanics every reference frame needs only one ideal clock. But a reference object, an origin and directions of coordinate axes can be chosen at will. It is well known how this circumstance is utilized in geometry. For example, the equation of an ellipse has the simple form $x^2/a^2 + y^2/b^2 = 1$ only if the origin is placed in the centre of the ellipse and the coordinate axes coincide with its principal axes. No doubt, all typical features of the ellipse remain for any other choice of the coordinate system, but all formulae become immeasurably more complex. It is important to point out here that in analytical geometry the transition from one coordinate system to another varies only the algebraic form of equations of geometric objects while the objects themselves naturally remain invariable.

Considering physical phenomena, one may also set up a coordinate system rather arbitrarily. However, the two most significant properties of vacuum space are implicitly meant in this case: uniformity and isotropy. Uniformity is identity of all points in space. This property is very essential. Actually, it enables us to use physics. Laws of physics prove to be the same at various points of the Earth, and everywhere within the solar system, for that matter. But this is just what permits the origin to be placed at any convenient point. When we turn a coordinate system around the origin, we do not expect anything to change. This implies that all directions running from a given point are identical in their properties. And this is exactly how isotropy of space is defined. In classical mechanics, or, more precisely, in reference frames where the Newtonian laws are valid (see § 1.5), uniformity and isotropy of free space are assumed.

In contrast to geometry, in physics there is another choice of reference frames: one may consider those moving relative to one another. This is quite superfluous in geometry. But in physics the reference frames moving relative to one another are the inevitable occurrence. For example, physical experiments can be carried out abord a spaceship and on the Earth. These are the two reference frames, each of which may have instruments motionless relative to the frame. As soon as we accept the reference frames moving relative to one another, the two intrinsically different but fundamental questions crop up.

1. How does the motion of a reference frame affect physical phenomena observed in it, i.e. do the physical laws change on transition from one such frame to another?

2. Suppose we observe a concrete physical phenomenon by means of instruments resting in a certain reference frame, and obtain some values as a result of measurements of physical quantities characterizing this phenomenon. The same phenomenon can be observed in another reference frame moving relative to the first one. The measurements conducted in the second coordinate system will give us certain numbers defining the same physical quantities. How do these quantities correlate?

It is important to note here that the same phenomenon is observed in both systems. We must know how to correlate these quantities. After all, a reference frame is an artificial construction created for measurement purposes. The phenomenon itself, just as the laws of nature, cannot be affected by the choice of a reference frame. The natural phenomenon is the objective reality existing outside our senses and measurements.

Of course, the results of measurements may prove to be different in different reference frames but in any case we must know how to convert the results of observations obtained in one frame into those that are obtained, or can be obtained, in another. In short, we need a method to transform results of measurements. How can such a method be found?

The answer to the first question leads us to the principle of relativity and via Newton's laws helps to distinguish a special class of reference frames, that is inertial frames (§ 1.5). The answer to the second question is given by the rules for the transformation of coordinates of an event, i.e. the Galilean transformation in classical mechanics (§ 1.3) or the Lorentz transformation in relativistic mechanics (§§ 2.4, 2.5, 2.7).

§ 1.3. The Galilean transformation. The transition from one reference frame to another one moving relative to the first one was performed long before the advent of the theory of relativity. Apparently the first to use the technique was Huygens who treated in this way the problem of the collision of spheres. For the sake of brevity we shall designate the reference frame by the letter K, and provided there are several frames we shall introduce superscripts (K^0, K', K'', ...). We have mentioned that an event can be considered as an "element" of some physical phenomenon. It is natural to begin with the conversion of the quantities characterizing an event when a transition from one reference frame to another takes place. From now on "the transition from one reference frame to another" will be everywhere understood as the consideration of those reference frames which move relative to one another. A shift of the origin as well as a rotation of coordinate axes will not be taken for a "transition".

In an arbitrary reference frame K an event is described by the four numbers: x, y, z, t. The first three of these are the coordinates

of the point at which the event happened, while the last one specifies the moment of time at which it happened. We want to know how the same four numbers x', y', z', t' look in another reference frame K' moving relative to the frame K.

From the very beginning we are compelled to restrict our problem and consider only those reference frames which move uniformly and rectilinearly relative to one another, and do not rotate around the origin. In other words, none of the considered reference frames moves with an acceleration relative to another frame. Somewhat later it will become clear that we deal here with the collection of so-called inertial frames of reference. However, since such reference frames can be discriminated only on the basis of Newton's laws, we shall postpone the definition and identification of such frames till § 1.5. For the present, we shall consider the two frames K and K' moving uniformly and rectilinearly (translationwise) relative to each other, in terms of geometry. Let us assume that the reference frame K' moves relative to K at a velocity $\boldsymbol{V}$. Suppose that at a given moment of time t the radius vector of the point M in the frame K' is equal to $\boldsymbol{r}'$. Then it can be seen from Fig. 1.1 that $\boldsymbol{r}' = \boldsymbol{r} - \boldsymbol{R}$, where $\boldsymbol{r}$ is the radius vector of the same point in the frame K, and $\boldsymbol{R}$ is the radius vector of the origin of the coordinate system K' taken from the origin of K. This relation is valid for any moment of time and $\boldsymbol{R}$ varies according to the familiar law $\boldsymbol{R} = \boldsymbol{V}t + \boldsymbol{R}_0$, where $\boldsymbol{R}_0$ is the radius vector specifying the location of the origin O' at the moment of time $t = 0$. Taking into account that at the moment $t = 0$ both origins coincide, $\boldsymbol{R} = \boldsymbol{V}t$, and we obtain the coordinate transformation law in the vector form:

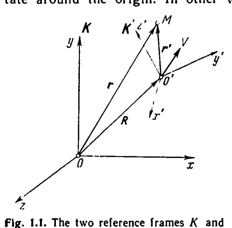


Fig. 1.1. The two reference frames K and K' with the arbitrarily directed axes x, y, z and x', y', z'. The frame K' moves relative to K at the velocity $\boldsymbol{V}$. The radius vector of the point M is equal to the vector $\boldsymbol{r}$ in the frame K and to $\boldsymbol{r}'$ in the frame K'. According to the vector summation rule $\boldsymbol{r} = \boldsymbol{r}' + \boldsymbol{R}$, where $\boldsymbol{R}$ is the radius vector of the origin O'.

$$\boldsymbol{r}' = \boldsymbol{r} - \boldsymbol{V}t, \tag{1.1}$$

where the components of the vector $\boldsymbol{V}$ are defined in the frame K. Now we can resort to isotropy of space and rotate each of the systems K and K' around its respective origin. It is convenient

to perform this in the following way. First, rotating the reference frames, we orient the x and x' axes along the direction of the relative velocity of these frames K and K'. Then rotating the frames around the common axis x, x', we orient axes y, y' and z, z' in parallel to each other. In such a way, having lost none of the generality in terms of physics, we come to the relative position of the coordinate systems shown in Fig. 1.2. In this case the velocity $\boldsymbol{V}$ has the components $(V, 0, 0)$. The origin of the system K' slips along the common axis at the velocity V, while at

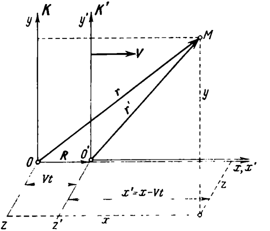


Fig. 1.2. The two reference frames K and K' with parallel axes move relative to each other at the velocity $\boldsymbol{V}$ ($\boldsymbol{V}$ is the velocity of motion of K' relative to K). In classical physics coordinates of an "event" are transformed from the reference frame K to K' according to the formulae of the "Galilean transformation": $x' = x - Vt$, $y' = y$, $z' = z$, $t' = t$.

the initial moment of time both origins coincided. It is seen from the vector formula (1.1) or directly from Fig. 1.2 that the relationship between the coordinates of the point M, where we believe the event happened, in the systems K and K' is determined by the following equations:

$$x' = x - Vt, \quad y' = y, \quad z' = z.$$

Now, in order to establish fully what are the coordinates of the event in the system K', one has to know the time of the event by the clock of the system K' (now we have two clocks: one in the system K and another in the system K'). Since in both systems we employ infinitely fast signals, the finite relative velocity of the systems is inessential to such signals. Indeed, the infinite velocity remains infinite in both systems. Consequently, the time of

the event registered by the clocks of both systems will be the same, i.e. $t = t'$. This conclusion is confirmed by our own "common sense", because we do not detect any influence of the motion on the clock rate in everyday life. But we should bear in mind, however, that infinitely fast signals were only assumed, and although the common sense does not deceive us in everyday life, we must be prepared to the fact that in the case of a finite velocity it may turn out that $t \neq t'$.

But within the scope of classical mechanics, as we have established by now, the formulae of transformation from the "coordinates" of an event determined in the system K (x, y, z, t) to those of the system K' (x', y', z', t') may be written as follows:

$$\begin{aligned} x' &= x - Vt, \\ y' &= y, \\ z' &= z, \\ t' &= t. \end{aligned} \tag{1.2}$$

Naturally, these equations are valid only for the relative position of the reference frames shown in Fig. 1.2. The transformation of the event "coordinates" from the frame K to the frame K', as given by Eqs. (1.2), is called the *Galilean transformation.* We would like at once to draw readers' attention to the fact that time turns out to be the fourth coordinate of an event so that when speaking of coordinates of an event, we imply four numbers (x, y, z, t). This is done not only for the sake of speech brevity. In the special theory of relativity such a terminology gains a complete justification (see Chapter 4).

We have already pointed out the equivalence of frames moving uniformly and rectilinearly relative to one another. The frames K and K' that will be referred to from now on differ only in the velocity of K' relative to K being equal to V, whereas the velocity of the frame K relative to K' is equal to $-V$. Hence, in order to get the reverse transformation formulae, it is sufficient to make primed and unprimed quantities change place, having changed the sign of V in the process. We get

$$\begin{aligned} x &= x' + Vt', \\ y &= y', \\ z &= z', \\ t &= t'. \end{aligned} \tag{1.3}$$

Surely, these same equations can be derived in a direct algebraic manner.

Note one of the consequences of the Galilean transformation. Suppose two events occurred in the frame K on the x axis: one

at the point x_1 at the moment t_1, and the other at the point x_2 at the moment t_2 ($t_1 \neq t_2$). Is it possible to select the frame K' in which both events would happen at the same point in space? Let us find the x coordinates of these events in the frame K': $x_1' = x_1 - Vt_1$, $x_2' = x_2 - Vt_2$; and compose the difference $x_2' - x_1' = x_2 - x_1 - V(t_2 - t_1)$. Having bidden $x_2' - x_1' = 0$, we obtain the equation from which the velocity of the frame K' relative to K is determined: $V = (x_2 - x_1)/(t_2 - t_1)$. The meaning of the result is very simple: during the time period $t_2 - t_1$ the frame K' succeeds in bringing the point x_1' to the place where the second event occurred by a requisite moment. We see that it is always possible to select a frame K' satisfying the required condition. It is possible, however, only because classical mechanics permits the velocity V to have any magnitude. In the theory of relativity, where the velocity of a reference frame, just as any other material object, is limited, the required frame is far from being always found.

Before proceeding to the Galilean principle of relativity, let us agree on one term. For the ease of speech "various observers" or "observers in different reference frames" are often mentioned. In the past such a terminology provoked blustering arguments, because there were many who imagined that it implied a subjective approach to physical measurements. But the presence of an observer is not at all mandatory as far as measurements are concerned: they can be taken by means of instruments and without man's assistance. It is indeed the case, for example, with spaceships, even when there are people aboard. "An observer from a frame K" is, in fact, taken to mean a set of instruments resting in this frame. One should not be surprised by the fact that instruments placed in different reference frames will give different results for measured quantities associated with one and the same phenomenon inasmuch as relative motion is a fundamental physical quality. Objectivity of laws of nature manifests itself when from results of observation in one reference frame one can find results of observation of the same phenomenon in any other frame. One may hope that after these remarks the appearance of an "observer" on the pages of this book will not give rise to any objections.

§ 1.4. The Galilean principle of relativity. Newton's second law. The Galilean principle of relativity pertains to mechanical phenomena exclusively; it was the first step toward the establishment of the principle of relativity that later embraced all physics. Galileo noticed that uniform and rectilinear motion does not affect mechanical phenomena. It is necessary to formulate precisely what it means. As we already know, a reference frame is needed to describe any physical phenomena, including mechanical ones. Let us consider two reference frames moving uniformly and recti-

linearly relative to each other and let us conduct a "mechanical experiment" in one of these frames. For example, we shall study the motion of a mathematical pendulum or the free fall of bodies. The principle of relativity states that identical experiments conducted in the two frames mentioned yield identical results. Hence, it is impossible to detect relative motion of the frames by means of such experiments. Of course, relative motion is easy to detect in many ways provided the experiments of another kind are undertaken. The first formulation of the principle of relativity can be found in Galileo's book *Dialogue Concerning the Two Chief World Systems — Ptolemaic and Copernican* (1632). This formulation is of purely qualitative nature. We shall quote a short extract from this book illustrating the essence of the problem:

"Shut yourself up with some friend of yours in the main cabin below decks on some large ship, and have with you there some flies, butterflies, and other small flying animals. Have a large bowl of water with fish in it: hang up a bottle that empties drop by drop into a wide vessel beneath it. With the ship standing still, observe carefully how the little animals fly with equal speed to all sides of the cabin. The fish swim indifferently in all directions; the drops fall into the vessel beneath; and, in throwing something to your friend, you need throw it no more strongly in one direction than another, the distances being equal; jumping with your feet together, you pass equal spaces in every direction. When you have observed all this carefully (though there is no doubt that when the ship is standing still everything must happen in this way), have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that. You will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still."

The significant consequences follow from the qualitative formulation of the Galilean principle of relativity, which is the identity of results of identical mechanical experiments conducted by two observers moving uniformly and rectilinearly relative to each other. Indeed, if the laws governing mechanical phenomena are known and all identical mechanical experiments produce the same result regardless of a reference frame chosen, the laws of mechanics must also be identical in such frames. In other words, the equations of mechanics must be the same in all reference frames moving uniformly and rectilinearly relative to each other.

Thus, the principal equations of mechanics written via coordinates and readings of a clock of its own reference frame must have the same form. At the same time it is clear that many quantities vary on transition from one reference frame to another. In-

deed, let us examine the motion of a particle * in the frame K. Usually it is defined as a time dependence of a radius vector $\boldsymbol{r} = \boldsymbol{r}(t)$. According to Eq. (1.1) the motion of the same particle in K' is defined by the variable radius vector $\boldsymbol{r}'(t) = \boldsymbol{r}(t) - \boldsymbol{V}t$. Differentiating both sides of the last equation with respect to t and taking into account that $d\boldsymbol{r}'/dt = \boldsymbol{v}'$, and $d\boldsymbol{r}/dt = \boldsymbol{v}$, we obtain

$$\boldsymbol{v}' = \boldsymbol{v} - \boldsymbol{V}. \tag{1.4}$$

Hence the velocity of the particle in the frames K and K' is different. The quantities that vary on transition from one coordinate system to another are called *relative*. Thus x coordinates and the velocity of a particle are relative quantities. Its acceleration, however, is the same in both frames, K and K'. This becomes evident immediately after differentiating Eq. (1.4).

$$\frac{d\boldsymbol{v}'}{dt} = \frac{d\boldsymbol{v}}{dt} \qquad (\boldsymbol{V} = \text{const}).$$

The fact that the acceleration of objects is the same for all observers in frames moving uniformly and rectilinearly relative to each other, is immediately evident. But this result makes it possible for us to understand the statement that "the equations have the same form in all reference frames". The fundamental equation of classical mechanics is that expressing the second law of Newton. This equation relates a force F acting on a body and the acceleration acquired by it due to the action of this force:

$$m\frac{d\boldsymbol{v}}{dt} = \boldsymbol{F}; \tag{1.5}$$

the factor m is called the *mass* of a body.

If the laws of mechanics in all reference frames moving uniformly and rectilinearly relative to one another are really the same, Eq. (1.5) has to retain its form in all reference frames of this kind. It is not difficult to see that this is really the case. We have already shown that acceleration is the same in all reference frames being investigated. But what happens to forces on transition from one reference frame to another? Suppose we investigate two objects: I and II. Let the force of their interaction depend on the distance between them, their relative velocity and time. But the Galilean transformation does not change any of these quantities. Indeed, let us write out the coordinates and ve-

* What is meant is a small object possessing a mass, but still so minute that there is no need to take into account its rotation. In mechanics, in this case, they speak of a mass point, but since we shall have to deal with points in space far too much, the term "particle" is preferable.

locities of objects I and II in the frames K and K', using the Galilean transformation:

	Coordinates and velocities of objects I and II in K	Coordinates in K'	Transformation of velocities
I	$x_1,\ y_1,\ z_1;\ v_1$	$x'_1=x_1-Vt,\ y'_1=y_1,\ z'_1=z_1$	$v'_1=v_1-V$
II	$x_2,\ y_2,\ z_2;\ v_2$	$x'_2=x_2-Vt,\ y'_2=y_2,\ z'_2=z_2$	$v'_2=v_2-V$

At once it becomes evident that $x_2 - x_1 = x'_2 - x'_1$, $y_2 - y_1 = y'_2 - y'_1$, $z_2 - z_1 = z'_2 - z'_1$. And in so far as the distance between the bodies is equal to

$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2} \quad \text{in the frame } K$$

and

$$\sqrt{(x'_2-x'_1)^2+(y'_2-y'_1)^2+(z'_2-z'_1)^2} \quad \text{in the frame } K',$$

it is clear that it remains constant on transition from K to K'. As to the relative velocity,

$$v'_2 - v'_1 = v_2 - v_1,$$

i.e. it remains permanent. In accordance with the Galilean transformation time is invariant: $t = t'$. Consequently, the forces dependent on the variables cited do not at all vary on transition from K to K'. But the forces considered in mechanics depend either on a distance (gravitational forces, forces of electric interaction, elastic forces) or on a relative velocity (friction forces). Hence, forces occurring in mechanics stay permanent under the Galilean transformation. Inasmuch as all the quantities appearing in Eq. (1.5), accelerations and forces, do not vary under the Galilean transformation, the fundamental equation of classical mechanics, the second law of Newton, relating forces and accelerations, has the same form in the frames K and K' and differs only in the designations of variables. (Surely, it is assumed that a mass is a constant quantity; the mass invariance is one of the basic postulates of classical mechanics.*) The equation describ-

* Note that in Newtonian mechanics motion of bodies of a variable mass can be examined, e. g. when jet propulsion or motion of a drop accompanied by condensation is studied. But in all these cases a body either donates substance to the environment or acquires from it. When mass variability is mentioned in the STR (see Supplement IV), it is implied that a mass of a body stays constant in the resting frame, i. e. there is no mass exchange between a body and its environment.

ing the second law of Newton in the frame K has the following form, provided the force depends on a distance and time:

$$m\frac{d^2\boldsymbol{r}}{dt^2}=\boldsymbol{F}(\boldsymbol{r}_{12}, t).$$

Accordingly, in the frame K'

$$m'\frac{d^2\boldsymbol{r}'}{dt'^2}=\boldsymbol{F}(\boldsymbol{r}'_{12}, t').$$

An equation which does not change in case of a transformation of variables appearing in it, i.e. an equation with its terms invariant, is called *invariant* with respect to a given transformation. Thus we have shown that the equation describing the second law of Newton is invariant with respect to the Galilean transformation.

Now we can formulate more precisely the terms on which "identical experiments produce identical results". Newton's Eq. (1.5) is an ordinary differential equation of the second order. Its solutions describe motion of the system. To make the solutions of Eq. (1.5) coincident in the frames K and K', i.e. to ensure the "identity" of motion, it is necessary for the initial conditions to coincide. The invariance of the basic equation of mechanics ensures that mechanical phenomena proceed alike in all reference frames moving uniformly relative to one another, only when the initial conditions coincide in these frames.

When the initial conditions for the same phenomenon differ in diverse reference frames, the phenomenon itself will look different. For example, while a raindrop falls down vertically from the viewpoint of an observer standing on a platform, the same raindrop will move along a parabola from the viewpoint of an observer in a train. (We suppose that a raindrop falls down with an acceleration.) However, the initial data in these frames were different. From the viewpoint of an observer in a train the raindrop had initially a horizontal component of a velocity. An observer standing on a platform had to assume that at the initial moment a raindrop had no horizontal component of a velocity whatsoever.

We have already mentioned above that in Newtonian mechanics an interaction between bodies is assumed to be transmitted infinitely fast. Now we are able to explain this in detail. "Interaction" between bodies is specified by forces. In classical mechanics forces are regarded as being dependent on distances between bodies. The same is assumed to be correct for bodies moving relative to one another. But a distance between two moving bodies has to be put down as

$$r_{12}=\sqrt{[x_2(t)-x_1(t)]^2+[y_2(t)-y_1(t)]^2+[z_2(t)-z_1(t)]^2}.$$

Assuming that an interaction, that is a force, is transmitted at a finite velocity, one cannot presume that the equation defining this force still incorporates r_{12}. If we want to find the force that is exerted on body *I* by body *II*, the position of body *II* should be registered not at the moment t but earlier by the time interval needed for the interaction to be transmitted from body *II* to body *I*. When this time lag is ignored, it means that the velocity at which the interactions are transmitted is assumed infinitely fast. It is precisely how problems are tackled in Newtonian mechanics.

The same thing happens when a potential energy is introduced. Having written down a central force for the interaction of two particles as usual in the form

$$\boldsymbol{F} = -\nabla U(|\boldsymbol{r}_1(t) - \boldsymbol{r}_2(t)|),$$

we explicitly ignore the time lag in the interaction transmission.

The instantaneous transmission of interactions, formerly referred to as a long-range action, appears amazing and obscure to us. The transmission of any signal, i.e. an impulse or energy capable of accomplishing some action, e.g. switching on a certain device, requires some time. As our experience teaches us, it is impossible to transmit a signal from "here" to another place ("there") instantaneously. Yet in Newton's time no other idea except the long-range action could emerge, as far as a transmission of interactions is concerned. The finite velocity of the transmission of interactions appeared together with the concept of a field which was introduced into the theory of electromagnetism by Maxwell. In Maxwell's theory the interaction of charges or currents is realized through a field to which an independent existence is attributed. It follows from the theory that a field propagates at a finite velocity. This means that the velocity of propagation of interaction is the same as that of the field. The propagation velocity of an electromagnetic field *in vacuo* plays a fundamental role in the theory of relativity. It is designated by the letter c and is approximately equal to $3 \cdot 10^8$ m/s. Since field variations are transmitted from point to point, field theories are referred to as short-range ones. As we shall see later, the theory of relativity rejects the long-range action as a matter of principle.

§ 1.5. Newton's laws and inertial frames of reference. The basic laws of mechanics, Newton's laws, make it possible for us to distinguish among all conceivable reference frames, the special class of frames in which not only laws of mechanics but also all other physical laws look particularly simple. These are the so-called inertial frames of reference. An *inertial frame of reference* is a frame (or rather frames, since it will turn out later that there are an infinite number of them) in which all three laws of Newton are valid.

We begin by showing how important the first law of Newton is for the discrimination of inertial frames of reference among all others. The first law of Newton, the law of inertia, claims that a body subjected to no forces moves due to inertia, i.e., uniformly and rectilinearly. Frequently one could hear, or even read in a textbook, that the first law is not an independent statement, but only a consequence of the second law.

Formally it is the case. The resultant of all forces acting on a body appears in the right-hand side of Eq. (1.5). The second law just claims that an acceleration acquired by a body is directly proportional to this resultant and inversely proportional to its mass. It follows from Eq. (1.5) that if the resultant of all forces is equal to zero, or there are no forces whatsoever, the body gains no acceleration. And if a body gains no acceleration, it either moves uniformly and rectilinearly or is at rest. It used to be concluded from this that the law of inertia could be obtained from the law of dynamics.

Then why was it necessary for Newton to formulate the law of inertia separately? It is doubtful that Newton did not realize that the law of inertia is a consequence of the law of dynamics. The problem is more complicated than it may seem at first sight Newton understood very well that neither Eq. (1.5) nor the law of inertia can be equally valid in all reference frames. It is not accidental that the definition of an inertial frame involves all three laws of Newton. Let us recollect the third law: to every action there is an equal and opposite reaction. This law emphasizes that all forces in Newtonian mechanics are intrinsically associated with an interaction between bodies.

Let us examine one useful, even though very plain example. Let a body be at rest in an inertial frame of reference K. Then according to the second law of Newton no forces act on this body. Without touching it let us consider it from the viewpoint of an observer moving relative to the frame K with an acceleration $\boldsymbol{a}$. This observer will note that the body in question moves relative to him with an acceleration $-\boldsymbol{a}$. If the second law of Newton were valid in his frame, he could say that the body experiences the force $-m\boldsymbol{a}$. But we know from the observer in an inertial frame of reference that there is no force acting on the body. Therefore, the second law of Newton is merely not valid in the reference frame moving relative to the inertial frame with an acceleration. Many readers have already realized, of course, that passing into the reference frame moving at an accelerating velocity, we detect "a force of inertia" which is not actually a force in Newtonian mechanics (see Supplement V). Since the laws of Newton are not valid in all reference frames, Newton had to point out that a certain reference frame ***was available*** in which all these laws were

valid. And the first law of Newton is, in fact, equivalent to this statement. This law postulates that an inertial frame of reference, i.e. a reference frame in which the law of inertia is valid, is available. In other words, one can find a reference frame in which a body that interacts with no other bodies, moves due to inertia, i. e. uniformly and rectilinearly.

The law of inertia represents a special case of the law of conservation of momentum. On the one hand, it is a consequence of the second and the third laws of Newton, and on the other hand, a consequence of the second law and the assumption about uniformity of space (equivalence of all its points), i.e. Newtonian mechanics assumes the uniformity of space in any inertial frame of reference.

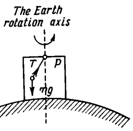


Fig. 1.3. The Foucault experiment designed to detect one of IFRs. For the sake of simplicity the drawing illustrates the Foucault experiment being performed at the Pole In fact, the experiment was conducted in Paris, but this circumstance does not change the matter.

Now suppose we have found one inertial frame of reference. Then according to the Galilean principle of relativity all reference frames moving uniformly and rectilinearly relative to it will be inertial as well. Therefore, it is clear that there is an infinite number of inertial frames of reference.

How does one find at least one inertial frame of reference? Of course, the discovery of such a frame is a matter of experience. The famous pendulum experiment first conducted by Foucault is suitable for the purpose. For the sake of simplicity we shall describe the experiment the way it could be conducted at one of the Earth's poles (Fig. 1.3). A heavy ball is suspended on a thread which is attached to a frame constructed at the Pole. The point of the pendulum suspension is located on the Earth's axis. The attachment of the thread is free and so the frame does not carry the thread along in the process of rotation around the Earth's axis. The equilibrium position of the pendulum thread coincides with the Earth's axis. If one deflects the pendulum from the equilibrium position and then lets it go without imparting an initial velocity, it will start oscillating in a certain plane. The two forces acting on the pendulum are the gravitational force mg and that of the tensile stress T of the thread. Both forces act in the plane P of the pendulum oscillations and cannot remove the pendulum from that plane. If the second law of Newton were strictly valid on the Earth, the plane of the pendulum oscillations would maintain its orientation relative to the Earth. But

in the experiment the Earth withdraws from under the pendulum and thereby "registers" the fact that in the coordinate system associated with the Earth the second law is not, strictly speaking, valid.

One should not be too much annoyed because of this, since Newton's laws still can be used on the Earth to great advantage. It is evident, for example, from the fact that the whole engineering and theoretical mechanics rely on the second law of Newton without any corrections. Surely, this is because corrections are small: they are caused by the Earth rotation which is not very fast. Therefore, the Earth can be treated as an inertial frame even in a school textbook.

But fundamentally the Earth is not an inertial frame. An inertial frame involves such a coordinate system relative to which the plane of pendulum oscillations remains constant. This plane can be found from the same Foucault experiment. The system turns out to be rather "exotic". Its centre is located in the Sun and the three coordinate axes are directed to the "stationary" stars, i.e. the stars moving rigidly together with the so-called celestial sphere. Due to the singular role of the Sun the inertial frame based on this system is referred to as *heliocentric*. In the choice of an inertial frame most important is the choice of directions for coordinate axes. The choice of the origin in the centre of inertia of the Sun is convenient because the Sun possesses the largest mass in the whole solar system. The motion of planets appears particularly plain in this frame. Note that the axes of the heliocentric reference frame do not participate in the rotation of the Sun. By the way, the reference frame with the coordinate axes fixed rigidly to the Earth, i.e. rotating with it, is referred to as *geocentric*. As Foucault's experiment showed, this frame is non-inertial.

Thus, the Newtonian laws of dynamics are applicable in the heliocentric frame. In accordance with the Galilean principle of relativity the laws of Newton are equally valid in all reference frames which move uniformly and rectilinearly relative to the heliocentric one. We shall refer to all these reference frames as inertial frames of reference *. Although the number of inertial frames of reference is infinite, they still get lost among all feasible kinds of frames. If it were possible to gather all kinds of frames into a sack and then to draw out of it one frame at random, we would get most likely a non-inertial frame.

Foucault's experiment is far from being the only one permitting of detecting a deviation of the geocentric reference frame from

* Hereinafter we shall often abbreviate the term "inertial frame of reference" to the initial letters, i. e IFR.

an inertial one. We shall indicate another experiment of this kind. When a heavy object is dropped from some height, it does not fall vertically down as it should due to the gravitational force, but deviates slightly to the east. The deviation of the motion of free falling objects from the vertical makes it possible to detect the non-inertial nature of the geocentric frame and to find an inertial frame of reference.

In mechanics there is one more conservation law for closed-type systems which is the conservation of moment of momentum. It is, just as the law of conservation of momentum in a closed-type system, the consequence of the second and third laws of Newton. Moreover, it can be obtained as a consequence of the second law and the assumption about the isotropy of space. This implies that Newtonian mechanics presupposes the uniformity of space.

The law of conservation of energy for closed-type systems turns out to be a consequence of the second law of Newton and the assumption about a potential character of forces acting between particles constituting the system. On the other hand, it stems from the motion equations of the system and the assumption about the uniformity of time. It follows that in Newtonian mechanics the uniformity of time is presupposed.

That is why an inertial frame of reference can be determined as one relative to which space is uniform and isotropic, and time uniform.

Sometimes an inertial frame of reference is defined as a frame fixed rigidly to a free-moving object. Although this definition is basically true, it cannot be practically used for the purpose of an experimental identification of an IFR. There is no "free-moving" object at our disposal, since the gravitational force cannot be cancelled. Consequently, it is more correct to define an IFR as a frame in which all three laws of Newton are valid.

Inertial frames are distinguished among other, non-inertial, reference frames not only in mechanics. An electric charge does not radiate electromagnetic waves when at rest in an inertial frame of reference, whereas in a non-inertial frame it does.

Inertial frames of reference play a tremendous role in physics. It is for these frames that familiar laws of physics are recorded. The transition to non-inertial frames is associated with considerable difficulties. The special theory of relativity instructs us how to describe all kinds of physical phenomena in any inertial frame of reference *. But what does it mean and how is it practi-

* It must be stressed that the STR can also be formulated for non-inertial frames of reference In fact, the STR can be employed in any reference frame, as long as there are no gravitational forces, i. e. in a plane four-dimensional space-time. However, the form of the STR suggested by Einstein and to be developed in this book is applicable only to inertial frames of reference.

cally carried out? We have much to discuss before we can get answers to these questions.

§ 1.6. Absolute time and absolute space. Although in deriving the Galilean transformation we have, in fact, already spoken of everything that was meant in that transformation by space and time, we shall repeat the pertinent statements. Usually, when "classical" physics is mentioned, Newtonian mechanics is implied for the views of Newton of space and time reflect precisely the classical approach to these concepts. Newton's ideas are worth dwelling on more carefully because they correspond to our everyday experience. and are customary and comprehensive, while the transition to the concept of space and time inherent in the special theory of relativity presupposes renouncing these ideas. In addition, a still more decisive step further from these concepts was made by Einstein in his theory of gravitation which is sometimes referred to as the general theory of relativity. This is what one can read in Newton's *Philosophiae Naturalis Principia Mathematica* (1687)· "Absolute space, in its own nature, without relation to anything external, remains always similar and immovable."

So, according to Newton, space represents a giant empty box which contains material objects and where physical phenomena take place. At the same time Newton was aware that the Galilean principle is valid in mechanics. And this indicates that the states of immobility and uniform rectilinear motion are equivalent. Then how should one single out "motionless absolute" space?

Of course, it is impossible to single out "motionless absolute" space just by observing mechanical phenomena. Detection of absolute space and absolute motion involves studies outside the scope of mechanics. Such a detection is assumed to be possible in the process of interpreting optical phenomena. Consequently, in the historic essay dedicated to the interpretation of some experimental facts (see Supplement II) we shall presume that Newton's privileged, selected reference frame, that is motionless absolute space, is the heliocentric frame. Finally, it will be clear, though, that there is no such thing as a privileged frame at all, but there is a whole privileged class of reference frames in which laws of physics appear particularly simple. Such is the class of inertial frames of reference.

Now let us see what Newton wrote about time:

"Absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external, and is otherwise called duration."

Again we come across the statement that time is something external relative to nature. Thus, in accordance with Newton's ideas, time and space exist by themselves and do not depend on material bodies located in space. Surely, Newton's concepts of

space and time seem very scholastic to us. However, they should not be underestimated. Here is a short excerpt from the book [11]:

"In conversations with one of the authors of this book at various times over the years, Einstein emphasized his great respect for Newton and, in particular, his admiration for Newton's courage. He stressed that Newton was even better aware than his 17th century critics of the difficulties with the ideas of absolute space and time. However, to postulate those ideas was the only practical way at that time to get on with the task of describing motion".

Of course, the natural question arises: why does classical mechanics based on such concepts of space and time that can hardly be explained, function so efficiently? It turns out, however, that these concepts are approximately correct and the departures from them in everyday life are quite insignificant. The departures from the classical ideas become clearly visible only when microparticles are investigated and also in outer space conditions which modern physics has already begun studying. Such investigations, however, require special conditions and sufficiently complex equipment.

To end this brief section it is necessary to give a concise presentation of the up-to-date approach to the problem. From the modern point of view there is no absolute space and, consequently, no absolute motion. All inertial frames of reference are equivalent. The special theory of relativity shows that time readings for events prove to be different in different inertial frames of reference. Thus, time reading is found to depend on the state of motion. The gravitational theory of Einstein goes still further. In terms of this theory properties of space and time are not prescribed for ever but are specified by objects located in space. Since in accordance with dialectic materialism space and time are forms of existence of matter, the conclusions of Einstein's theory of gravitation appear far more satisfactory than the Newtonian concepts of space and time.

§ 1.7. How physics was approaching the theory of relativity. From the point of view of modern physics it is useful to trace how relativistic effects were showing well before the creation of the special theory of relativity. This section does not claim to be an historic essay (Supplement II is closer to that). It is intended only for promoting the understanding of the next two sections, where, in fact, the first principles of the theory are presented.

No doubt, the first step to the development of the special theory of relativity was the discovery by Galileo of the principle of relativity for mechanical phenomena.

The natural question arises: why did Galileo confine his prin-

ciple within the framework of mechanics? The answer is very straightforward: in Galileo's times there were just no "other branches of physics" as we call them now. In fact, mechanics represented the whole physics. If one also takes into consideration that all physical phenomena were attempted to be explained on the basis of mechanics almost till the end of the 19th century, it becomes clear that the principle of relativity formulated by Galileo encompassed the "whole physics" at that time.

The next important step along the road to the theory of relativity was the establishment of the finiteness of the velocity of light. The conclusion was made by Roemer on the basis of his astronomical observations (1676). Before Roemer the velocity of light propagation was assumed to be infinite.

The Galilean principle of relativity could be expressed in a mathematical form only after equations of mechanics had been written down (Newton, *Philosophiae Naturalis Principia Mathematica*, 1687). Since coordinates and time are the basic variables involved in equations of mechanics, their transformation on transition from one reference frame to another, moving relative to the former, requires appropriate equations for transformation of coordinates and time under such a transition. It followed from the Galilean principle of relativity that the requisite transformation of coordinates and time should not alter the form of Newton's laws (§ 1.4). This transformation is that of Galileo.

In 1851 the Foucault pendulum experiment was performed at the Pantheon in Paris, which definitely demonstrated the Earth's rotation and indicated the inertial frame of reference (§ 1.5). In fact, one could conclude the description of mechanical phenomena, linked directly with the theory of relativity, with this experiment.

The Galilean principle of relativity, Newton's laws and the Galilean transformation are all closely interrelated. The direct consequence of the Galilean transformation is the classical formula for the velocity transformation (Eq. (1.4)): $\boldsymbol{v}' = \boldsymbol{v} - \boldsymbol{V}$. In 1851 Fizeau performed an experiment to show explicitly that this formula is not always correct. The Fizeau experiment with flowing water was schematically conducted as follows. In the reference frame K water was flowing along a tube at a velocity V, and the velocity of light in water was being measured. Now we can reason strictly in terms of kinematics. Let us fix the inertial frame K' to moving water. In this frame the velocity of light v' is determined by the familiar relationship $v' = c/n$, where n is the refraction index of water. To find the velocity of light in the frame K, one can use Eq. (1.4), and then $v = c/n + V$. But Fizeau's result, confirmed also by modern measurements, turned out to be

$$v = \frac{c}{n} + V\left(1 - \frac{1}{n^2}\right).$$

As it is seen from here, the classical Eq. (1.4) is not correct in this case. And this is exactly what we wanted to emphasize. As to the details of the experiment and its contemporary interpretation, all that can be found in § 3.6.

The theory of relativity owes very much to the Maxwell theory formulated first in several large articles and later published in the two-volume *Treatise* (1856-1873). That was the first field theory in which an interaction was supposed to be transmitted at a finite velocity, that is the field propagation velocity. The theory provided a quite definite value for this velocity which was specifically equal to $1/\sqrt{\varepsilon_0\mu_0}$ *in vacuo*, where ε_0 and μ_0 are the electric and magnetic constants. Naturally, the question arose at once as to whether the principle of relativity was satisfied; in other words, whether the Maxwell equations retained their form under the Galilean transformation. One can easily check that the Galilean transformation changes the appearance of the Maxwell equations. Owing to this fact it was suspected that the principle of relativity did not extend to dynamics. Several decades were needed to realize what wonder of a theory Maxwell developed. Neither knowing, nor even suspecting anything about the theory of relativity, Maxwell nevertheless developed his theory in a complete agreement with the requirements of the theory of relativity.

Now, when we know full well where the influence of the theory of relativity can be "detected", it is easy to come back to essential facts. The theory of relativity reveals itself when velocities of objects get closer to that of light *in vacuo*, such velocities being referred to as relativistic. However, there are no macroscopic objects possessing relativistic velocities. Only microscopic particles can travel at velocities close to that of light. The first microparticle to be discovered was an electron (Thomson, 1894-1896). Thomson determined the ratio of the charge of an electron to its mass experimentally. His experiments were carried out in discharge tubes where electron velocities were far below that of light. But in 1896 the natural radioactivity was discovered. The stream of electrons found among radiations emitted by radioactive substances was very soon identified with electrons in a discharge tube. Velocities of these electrons turned out to be close to that of light. When in 1902 Kaufmann investigated the motion of such electrons in electric and magnetic fields, the classical equation of motion, i.e. the second law of Newton, was found to describe their behaviour incorrectly. Thus, the departure from the Newtonian laws was observed for the first time.

Summing up, it can be said that by the beginning of the 20th century it became obvious that Newtonian mechanics and the Galilean transformation are not always true, and the fastest signals

of all known, that is light signals, are transmitted at a finite velocity.

Although there are no macroscopic objects moving at a relativistic velocity, one relativistic object, light, was always at men's disposal. Naturally, optical experiments played a significant role in the history of the STR: the interpretation of optical experiments is associated with the emergence of a hypothesis concerning a "luminoferous medium". Rejection of this hypothesis took much effort, but now it is worth mentioning only as a page in the history of physics (see Supplement II).

§ 1.8. The generalization of the Galilean principle of relativity. The Galilean principle of relativity covered only mechanical phenomena. We found that the second law of Newton expressed in a differential form in combination with the Galilean transformation satisfied the principle of relativity. From the formal point of view it implied that Eq. (1.5) remained invariant and only designations of variables changed. Naturally, the question arises: why must the principle of relativity cover only mechanical phenomena? Why is it impossible to believe that all physical phenomena happen in the identical manner in all inertial frames, provided the initial conditions of these phenomena are identically specified? In other words, why is it impossible to assume all inertial frames of reference to be completely equal with respect to all physical phenomena?

These questions did not worry physicists too much till the middle of the 19th century since they reduced all physics to mechanics. But by the middle of the 19th century it became evident that physics cannot be reduced to mechanics. By the same time the conviction had grown as to the universal relationship between phenomena, and between physical phenomena in particular. The subdivision of physics into "mechanics", "electricity", "heat" etc. is justified by the fact that each group of phenomena possesses its own set of basic equations and so is caused by rather educational requirements and is not intrinsically imperative. Looking more carefully into even "purely mechanical" phenomena, one can discern a manifestation of regularities of another kind. The collision of billiard balls is always cited as a classical example from mechanics. But at the moment of collision, when the balls are slightly flattened, the elastic forces defined by electromagnetic forces come into play. Hence, no "purely mechanical" phenomena can exist in nature. It follows that the principle of relativity must either cover "all physics" or be wholly incorrect.

Thus, the extension of the principle of relativity to all physical phenomena was quite natural from the viewpoint of physics at the end of the 19th century. But such generalization of the Galilean

principle of relativity is exactly what is called the first postulate of Einstein, or the Einstein principle of relativity.

However, the equations of electrodynamics were at once found to contradict the equivalence of inertial frames of reference.

First of all, so far as the basic system of electrodynamic equations, that is the Maxwell equations, is concerned, they alter their appearance under the Galilean transformation, i.e. do not retain their form, and it follows from here that electromagnetic phenomena are described differently in different IFRs. In other words, electromagnetic phenomena do not obey the principle of relativity. In particular, this means that in the reference frame in which the Maxwell equations are written down in the conventional form (see Chapter 6), the propagation velocity of electromagnetic waves c is equal to $1/\sqrt{\varepsilon_0\mu_0}$, while in all other reference frames moving relative to the first one, the velocity is different. But vacuum occupies a special place relative to reference frames. Indeed, it is remarkable because it has no "medium" possessing the rest mass. One can always fix a reference frame to a material medium, i.e. single out such a frame in which the medium is at rest as a whole or in a limited region. But this particular reference frame is the chosen one. Another equivalent frame of reference moving relative to the first one must possess the same property of the motionless medium. But this creates a different physical situation. Thus, the presence of a medium always distinguishes one reference frame from all others. But it is impossible to single out such a system *in vacuo* because there is no reference frame in which vacuum is at rest. Consequently, all reference frames are equivalent relative to vacuum. It follows logically from here that provided all inertial observers are equal, the velocity of electromagnetic waves must be the same, $1/\sqrt{\varepsilon_0\mu_0}$, in all IFRs.

As to the classical formula for the transformation of velocities, Eq. (1.4) shows that this is not the case. Let in an inertial frame of reference K the velocity of light *in vacuo* be equal to c. Then in another inertial frame of reference K' the velocity of light *in vacuo* c' is equal to $c - V$. Hence, the velocity of light *in vacuo* c is equal to $1/\sqrt{\varepsilon_0\mu_0}$ only in one privileged reference frame. Thus, the principle of relativity seemed to be incorrect for electromagnetic phenomena.

The above reasoning was based on the fundamental assumption which was absolutely unacceptable for the 19th century physics: electromagnetic waves, i.e. light, can propagate *in vacuo* or, expressed otherwise, no matter is needed for their propagation. This is a very difficult point to comprehend, when a transition from classical physics to relativistic is undertaken.

But what could be done in such a situation? Logically three possibilities were opening.

(1) The principle of relativity could be assumed to cover only mechanics and have nothing to do with electrodynamics in which there is an "absolute" frame of reference. But, as it was mentioned before, such a possibility is rejected when the general relationship of physical phenomena is taken into consideration.

(2) The principle of relativity could be regarded to be universally applied, and inasmuch as the system of Maxwell's equations does not satisfy this principle, that is it changes its appearance under the Galilean transformation, it should be discarded. But the system of Maxwell's equations showed itself as a reliable and comprehensive theory within one inertial frame of reference, a laboratory frame. On the other hand, Newtonian mechanics and the Galilean transformation associated with it did not prove to be always correct. Because of this it would be reasonable to keep the system of Maxwell's equations.

(3) If the principle of relativity is assumed to be applicable to all phenomena of nature, and the system of Maxwell's equations correct, the transition from one inertial frame of reference to another cannot be described by the Galilean transformation which changes the form of Maxwell's equations. On the other hand, a new transformation cannot leave the form of equations of mechanics intact. Consequently, the equations of mechanics have to be changed so that the new transformation leaves them intact.

The last possibility formulates concisely the programme which is realized by the special theory of relativity: (1) the principle of relativity covers all phenomena of nature, (2) the velocity of electromagnetic waves *in vacuo* is the same in all IFRs (this follows from the invariance of Maxwell's equations).

But how must a transformation of coordinates and time look like in order to meet both requirements set above? Such a transformation will turn out to be the Lorentz transformation and we shall examine it closely in the next chapter. In conclusion, we shall point out the following.

As soon as the Galilean principle of relativity was extended to cover all physical phenomena, it turned into a genuine principle of physics. Evidently it is advisable to differentiate laws and principles of physics. When laws of physics are spoken of, their validity for a limited scope of physical phenomena is implied. For example, Newton's laws describe phenomena of mechanics. Maxwell's equations pertain to electrodynamics, and so they are the laws of electrodynamics. The three laws of thermodynamics deal with thermal phenomena. As to the principles of physics, they are universally important, for they cover all physical phenomena.

The most widely known principle of physics is that of conservation of energy. The famous book by M. Planck dedicated to conservation of energy is called *The Principle of Conservation of Energy* (1931). We believe that the law of conservation of energy is true for all physical phenomena, just as we are sure that the law of conservation of momentum is true for all physical phenomena. The principle of relativity occupies its place in physics along with the principles of consetvation of energy and momentum.

§ **1.9. The velocity of light *in vacuo*.** The velocity of light *in vacuo* occupies a special place in nature because in accordance with present-day conceptions it is the greatest possible velocity at which an interaction between objects can be transmitted. Transmission of an interaction, i.e. transmission of a certain action produced by one object onto another, is often referred to as transmission of a signal; it is this term that is especially popular in the theory of relativity. To transmit a signal means to transmit a momentum and energy (taken to be inseparable in the theory of relativity (see § 5.5)) which are capable of "switching on" a certain device, e.g. a trigger mechanism.

It does not follow from anywhere that there exists an upper limit for the velocity at which signals can be transmitted in nature. However, both theory and experiment show that all known interactions propagate at a finite velocity; and the fastest velocity at which a signal is transmitted is that of light *in vacuo*. We shall recall that this is also the propagation velocity of electromagnetic waves of any frequency *in vacuo*. As was already mentioned, the classical theory assumed tacitly that a signal can propagate infinitely fast.

If one admits that there is an ultimate velocity of signal propagation in nature, its absolute value must be the same in all inertial frames of reference. In fact, all these frames are equivalent according to the principle of relativity, and it is impossible to suggest a physical experiment to detect the difference between them. Had the velocity of interaction transmission been different in different inertial frames of reference, it would have been possible to distinguish one inertial frame from another. This is impossible, however, provided the principle of relativity is assumed to be universal. It follows immediately from this that the velocity of light *in vacuo* must be the same in all inertial frames of reference.

And what if a source moves toward an observer or an observer moves toward a source? Such a motion cannot change the magnitude of the ultimate velocity at which a signal is transmitted. Consequently, the velocity of light *in vacuo* cannot depend on the motion of either a source or an observer.

Obviously, the velocity of light *in vacuo* has unique properties. All velocities are relative, i.e. they change on transition from one inertial frame to another. But the absolute value of the velocity c remains the same. Although there is no privileged frame among all inertial frames, there is one privileged velocity in all of them. Both these circumstances are intrinsically associated with the fact that electromagnetic waves can propagate *in vacuo*. In other words, no material medium is needed for their propagation. Naturally, the assumption concerning the privileged, i.e. invariant, velocity upsets drastically the classical arrangement, that is Eq. (1.4) and, hence, Eqs. (1.2) and (1.3).

Nevertheless, no matter how strict or beautiful the logical reasoning is, an experiment was and will for ever be a supreme judge in physics. An experiment supports quite unambiguously the following two statements: (1) in a given IFR the velocity of light *in vacuo* is equal at all points and in all directions; (2) in all IFRs this velocity has the same value. Here we refer to the Michelson-Morley and Kennedy-Thorndike experiments described in Supplement II.

CHAPTER 2

THE EINSTEIN POSTULATES. THE INTERVAL BETWEEN EVENTS. THE LORENTZ TRANSFORMATION

§ **2.1. Einstein's postulates.** The final part of the foregoing chapter was devoted to the explanation of two basic assumptions of the STR which are called the Einstein postulates. On account of their significance we shall repeat them once more here and supplement some comments.

Postulate I. *All identical physical phenomena proceed alike in inertial frames of reference in the case of equal initial conditions.* In other words, there is no privileged frame among IFRs, and the state of absolute motion is impossible to find.

This postulate extends the Galilean principle of relativity to all phenomena of nature. It puts an end to absolute space once and for all: since all inertial frames of reference are equivalent, they cannot have any privileged frame among them. It was just absolute space that served as such a privileged frame. The conception of "absolute" motion *in vacuo* which was meant as the motion relative to the absolute frame of reference is rejected exactly in the same way (see § 1.6).

Postulate II. *The velocity of light in vacuo is equal in all directions and in any region of a given inertial frame of reference, and equal in all inertial frames of reference.*

Often this postulate is supplemented with the statement that the velocity of light *in vacuo* is not affected by the velocity of a source. This, however, follows immediately from Postulate II formulated in the form given above. Indeed, any source can have an inertial frame of reference fixed rigidly to it. When a source moves non-uniformly and/or along a curved line, an instantaneous co-moving inertial frame can be found. In such a frame a source is at rest, while all other inertial frames move relative to it (and a source moves relative to them). Since in accordance with Postulate II the velocity of light is the same in all frames, it does not depend on the velocity of a source. As to the motion of the observer, it is the relative velocity of a source and an observer that is essential, so that the preceding reasoning disposes of the question.

It should be clearly realized what Postulate II implies. For this purpose let us imagine that the velocity of light is measured in the

frame K in the following way. At the moment t_1 a light signal is sent from the point x_1 along the x axis. It reaches the point x_2 at the moment t_2. Then $c = (x_2 - x_1)/(t_2 - t_1)$. Now the same two events, that is the sending and reception of a signal, are viewed from the frame K'. The sending of a signal occurs for an observer from the frame K' at the point x_1' at the moment t_1' and the reception at the point x_2' at the moment t_2'. In spite of the fact that the frames K and K' move relative to each other along the common axis x, x', we have to get the ratio $(x_2' - x_1')/(t_2' - t_1')$ equal to c. From the viewpoint of the "common sense" this must not be the case. (This becomes clear if one draws the diagram of the experiment.) However, this is exactly what Postulate II prescribes.

We have formulated Postulate II, in fact, the way it was done by Einstein himself in his article of 1905. However, in our time it is advisable to formulate it otherwise, namely, to proceed from the assumption that there is the ultimate velocity of signal transmission in nature. This is the principal assumption. Then this ultimate velocity is identified with the velocity of electromagnetic waves, i.e. light, *in vacuo*. The last assumption is not obligatory: basically, the STR would not have lost its meaning if the ultimate velocity had turned out to be different. However, the STR makes use of just this assumption. If one assumes that the velocity of light *in vacuo* is the ultimate velocity at which an interaction can be transmitted, it follows directly that it must have the same magnitude in all IFRs (see § 1.9).

Having formulated the first principles of the theory of relativity, that is two Einstein's postulates, one can formulate the general objective of the special theory of relativity. Its basis is the principle of relativity, i.e. the equivalence of all inertial frames of reference with respect to all physical phenomena. The theory of relativity has to give such a description of physical phenomena which will be the same in all inertial frames of reference. Thus, if we have some equations at our disposal describing one or another group of phenomena, these equations must appear alike in all inertial frames of reference, each frame using its own variables. Recall that equations of mechanics and electrodynamics intrinsically contain coordinates of an event and a moment of its occurrence. These coordinates and the moment of time registered for an event are transformed on transition from one inertial frame of reference to another. The Galilean transformation changes the appearance of Maxwell's equations, but since we want to preserve them as equations of an electromagnetic field correct in all inertial frames, we ought to find such a transformation of coordinates and time that will keep the appearance of Maxwell's equations invariable. Such a transformation will turn out to be the Lorentz transformation.

The Lorentz transformation, however, follows directly from the Einstein postulates. The point is that the Maxwell theory was developed from the very beginning as the relativistic one. The inherent cause for this consists in the fact that it described correctly the properties of light, the most relativistic object of all.

Thus, having found the transformation of coordinates and time satisfying the Einstein postulates, we have to be sure that the basic equations of physics are the same in all inertial frames, i.e. covariant relative to this transformation. The meaning of the term "covariant" will be explained in § 4.3. Now we have to dwell on the "basic laws" of physics.

The laws of Newton are referred to as the basic laws in mechanics, the laws of Maxwell as the basic laws in electrodynamics, and the equations expressing the first and the second principles as the basic laws in thermodynamics.

Relative quantities were known in classical physics, e.g. velocities, coordinates, velocity directions, but the special theory of relativity adds to them, rather unexpectedly for our intuition, relative time intervals between events and relative scale lengths, i.e. distances. However, this is the "price" that we must pay in order to realize the principle of relativity with respect to all physical phenomena.

And still the predominant feature of the theory of relativity, in spite of its title, is not at all the relativity of various quantities, i.e. their dependence on the choice of an inertial frame of reference. The essence of the theory of relativity consists in just the opposite. The theory of relativity shows that the laws of nature in inertial frames of reference do not depend on the choice of a reference frame and on a position and motion of an observer, but measurement results in different reference frames can be correlated. Speaking in terms of philosophy, the theory of relativity underscores the objective character of the laws of nature and not the relativity of knowledge.

Of course, trying to alter an historically established name, which by the way was proposed not by Einstein but by Planck in 1906, is a hopeless undertaking. However, there is one point to pay attention to. The controversy over the correct name for the theory, "special" or "partial", is not essential. In a sense, the problem is how to restrict the theory to inertial frames of reference. Essentially this restriction results in the theory which is correct in the absence of gravitational fields or, practically, in weak gravitational fields. That is why the most correct name would be the "restricted" theory of relativity, which is adopted in the French literature.

Although the Einstein postulates are the first principles of the theory of relativity, they are not sufficient for its development.

The construction of a relativistic frame of reference is fundamentally important for the theory, so we turn our attention to this aspect now.

§ 2.2. The relativistic frame of reference. In the construction of a relativistic frame of reference, just as in the construction of the whole theory, the validity of both Einstein's postulates is assumed. Besides, we shall assume that the velocity of light *in vacuo* is the ultimate velocity at which signals are transmitted. The last assumption is not present in the Einstein postulates. However, as we shall see later on, it has to be inevitably incorporated in the theory if we want the principle of causality to come into effect (see § 3.4). In § 1.1 we spoke in detail on how a reference frame is constructed in classical mechanics. There we indicated that it was sufficient for each reference frame to have one clock, since it was assumed that infinitely fast signals could be used. But in the STR the existence of a finite velocity of a signal is explicitly allowed for, so that when the velocities in question get closer to that ultimate one, it becomes inconvenient and even impossible to use only one clock. But it is just these velocities that are of interest for the theory of relativity.

Therefore, a set of clocks is to be added to a coordinate system constructed exactly in the way described in § 1.1. Basically, the STR implies that clocks are located at every point of space. This is not needed in practice, but as a matter of principle a clock must be at any point where the moment of an event is registered. All the ~~clocks of~~ ~~a given~~ reference frame are motionless relative to it.

It is assumed in the STR that it is possible to have at one's disposal as many ideal identical clocks as one needs. This assumption is easily realized in our time. According to quantum mechanics all microparticles of the same kind are identical. In particular, characteristic oscillation frequencies of atoms of the same kind coincide precisely. Taking the atoms themselves for the clocks and the periods of atomic characteristic oscillations for the time standards, we obtain a sufficient number of required clocks.

Length standards can be dealt with in just the same manner. The wavelength of a characteristic radiation of a given atom can be chosen quite adequately as a length unit. Even prior to the advent of quantum mechanics it was believed that the wavelength of a radiation of a given atom can be utilized as an invariable length standard: to wit, that is how the length of the metre was immortalized by Michelson at the beginning of the 20th century.

When we consider two IFRs moving relative to each other, the length scales and the clocks of each frame are at rest only with respect to "their own" reference frame. Is it possible to believe that we have identical length scales and clocks in different IFRs,

if, say, there are such scales and clocks only in one frame? In some literature one can come across a discourse to the point that length scales and clocks can be transferred from one IFR into another. There is no doubt that one should not do this. Transferring clocks and length scales from one IFR into another, we impart acceleration to them. Theoretically, acceleration varies the length of scales and the clock rate. Here are some straightforward examples: drop a clock or a ruler on a stone floor. A clock may just stop, and a ruler break. Even an atomic clock breaks down when atoms get destroyed. All that is the effect of acceleration.

But in order to obtain identical length and time standards in different IFRs, one does not need to transfer anything from one frame into another. It is sufficient to take a pure substance in any reference frame and its radiation will provide us with required standards. It should be emphasized how important it is to have length and time standards in each IFR which are truly identical with those in all other frames. Indeed, the principle of relativity and the equivalence of all IFRs in combination with the identity of length and time standards make it possible to attain the complete identity of these reference frames.

So, every IFR has as many adequate clocks as needed. The time of an event at a given point is the reading of the clock located at the point where the event occurred at the moment of the occurrence of the event. If two events occurred at different points in space and the clocks at these points registered the same time for the occurrence of these events, we have to regard these events as simultaneous. But obviously the synchronism of events occurring at different points in space depends on how the initial time readings of these clocks were adjusted, the clock rates being assumed absolutely identical. Thus the determination of the synchronism of events and the adjustment of the initial time readings of all the clocks belonging to a given IFR, i.e. the clock synchronization, are the same thing. It should be pointed out that the clock synchronization, that is the determination of the synchronism of events, can be accomplished in different ways. The advantages of the synchronization suggested by Einstein will be explained later on. All the same, it should be emphasized that the synchronism of events *is determined*, and this determination can be accomplished not in a single way.

Here is the example showing how important it is to know how to determine the synchronism of events. How is a velocity of a particle found? Let a particle move along the x axis. To obtain its velocity, one must know the position x_1 of the particle at the moment t_1 and also its position x_2 at the moment t_2. Provided the motion is uniform the velocity is equal to $(x_2 - x_1)/(t_2 - t_1)$. But the arrival of the particle at the point x_1 is registered by a

clock located at that point and the arrival at the point x_2 by a clock located at the point x_2. To determine the velocity, one has to be sure that the clock located at the point x_2 was showing at the moment t_1 the same time as the clock located at the point x_1. Only in that case the determination of the velocity would have any sense. But this just implies that the clocks must be synchronized.

Having explained that the determination of synchronism and clock synchronization are the same thing, we pass over to the procedure of clock synchronization within one IFR. The first thing to come to one's mind is to collect the clocks at one spot, verify them and then return them to their respective points. Following Einstein we shall reject this procedure, because every clock transfer is associated with an acceleration that the clocks gain. Theoretically, every acceleration affects a clock rate. Consequently, it is better first to set the clocks at their respective points and only then to verify them *.

How can one verify, i.e. synchronize, the clocks located at various points in space? Let a clock which we shall call a reference one be located at the origin of a given IFR. Of course, this particular clock does not differ in any detail from all others. One can send a signal from the reference clock to any clock of a given IFR. It is assumed in this case that the distances from each of the clocks to the reference one are known, with no clocks being necessary to determine these distances. Knowing the velocity at which a signal is transmitted, one can find the time it takes a signal to travel from the reference clock to any clock of the frame. If the signal from the reference clock is sent at the moment $t=0$, a synchronized clock should display just exactly this time at the moment when the signal reaches this clock. Although generally speaking one can use any signal, it is most convenient to choose a light signal *in vacuo* for the purpose of clock synchronization in all IFRs, since it propagates at the same velocity in all IFRs. The utilization of a light signal *in vacuo* for clock synchronization is one more factor ensuring the complete equivalence of all IFRs.

Thus, a synchronization "agent" is a light signal. Let us describe now a synchronization procedure for a given IFR according to Einstein.

1. Clocks are set at their respective points and actuated. Coordinates of points at which the clocks are located are known, and

---

* There exists the voluminous literature "illustrating" that the infinitely slow transportation of clocks does not affect their rates. No doubt, it is plausible in terms of physics. But inasmuch as the relativistic theory deals with relativistic velocities and distances, a procedure of this kind is hardly of any interest to us.

so the distances from each of the clocks to the reference one are known.

2. At an arbitrarily chosen moment t_1 a light signal is sent from the reference clock to the clock to be synchronized. The light signal travels *in vacuo* along the known path and its arrival is registered by an observer or by means of a device.

3a. The reading of the clock at the moment of the signal arrival is to be set to $t = t_1 + r/c$, where r is the distance to the reference clock. The "initial" reading of the clock is thereby chosen, the clock is "verified" against the reference one.

One may also use another equivalent method.

3b. Mirrors are set at all points where clocks are located in order to reflect light back to its source. If the reference clock registers the return of the signal at the moment t_2, the moment registered by the clock at the mirror is to be set equal to $t = t_1 + (t_1 + t_2)/2$.

The last procedure of clock synchronization has one delicate point. Using procedure 3a we presume the velocity of light c known. But we have already seen that two synchronized clocks are necessary to determine the velocity of motion in one direction. On the other hand, the velocity of light is usually determined from the motion of a beam along a closed path. In particular, the velocity of light could be found by means of reflection from a mirror, when only one clock is available. In this case one has to make use of procedure 3b and to know the distance from the reference clock to the mirror. If this distance is equal to r, then $c = 2r/(t_2 - t_1)$. However, if the velocity of light propagating "there" is not equal to that of light propagating "back", we are not able to establish this fact. It is impossible to ascertain this fact experimentally just because our clocks are synchronized to give the value c for the velocity of light. The theory of relativity, however, proceeds from the assumption that the velocity of light *in vacuo* is the same in all directions. Besides, the totality of experimental data does not contradict either this statement or the consequences of the theory of relativity.

Thus, we have come to a relativistic frame of reference comprising a coordinate system of rigid axes and synchronized clocks fixed rigidly to this system. Such clocks in a given IFR will be referred to as a "set" of clocks. The synchronization procedure according to Einstein is such that it can be performed in the same manner in any IFR.

In accordance with the adopted rule for clock synchronization, synchronism of events can also be determined as follows. Let two events occur at the points of space equally removed from the third point. If at the moment of the occurrence of the two events the light signals are sent from the points of events to that third

point, the events are assumed simultaneous if both signals reach the third point at the same moment of time.

Of course, accelerated motion of bodies can be treated in the framework of the STR, whereas accelerated motion of reference frames, relative to inertial ones, cannot be considered. Since length standards and clocks are rigidly fixed to their IFR, it is clear that these standards and clocks should not be accelerated. Otherwise, the study of an influence of acceleration on length standards and clocks would make us examine their specific structure and deprive the theory of its universal character.

§ 2.3. The direct consequences of Einstein's postulates (a few imaginary experiments). The two direct consequences of Einstein's postulates, "the relativity of length scales" and "the relativity of time intervals between events", can be obtained directly from the postulates themselves. Most often they are obtained from a transformation of coordinates and time of an event. This transformation is compatible with the Einstein postulates and is called the Lorentz transformation. However, this convenient method to be discussed in § 3.2 is not at all obligatory. Now we shall describe a few "imaginary experiments" by means of which we shall draw necessary conclusions. Imaginary experiments play a conspicuous role in conclusions of the STR. They represent some hypothetical experiments not to be necessarily conducted in practice. In fact these are only generalizations permitting definite consequences to be obtained from the given premises.* Now we pass over to a description of several imaginary experiments whose results we shall obtain once more when consequences of the Lorentz transformation are discussed.

We shall begin with a very simple imaginary experiment which illustrates the relativity of synchronism, provided the second Einstein postulate is satisfied. Later on we shall obtain the same result by different methods. The experiment is performed in the Einstein train. This term is applied to any train moving uniformly and rectilinearly at, preferably, a relativistic velocity. In an imaginary experiment one can assume even such a thing. The middle of the train is easy to find precisely. This is done in the train's reference frame and does not present any difficulties. Observer 1 is located in the middle of the train and observer 2 at the station. Light signals are sent to observer 1 from the ends of the train which are equally removed from him. The imaginary experiment is so performed that the signals travelling from the

* One should not think that "imaginary experiments" (Gedankenexperimente) are characteristic only of the theory of relativity. The series of "imaginary experiments" devoted to quantum mechanics can be found in the discussion of N. Bohr and A. Einstein, UFN **66**, 571 (1958).

ends of the train reach observer 1 just at the moment when he rides up to observer 2. In the imaginary experiments one does not usually take interest in how to accomplish this practically. It is essential to us what conclusions will be drawn by the two observers from the fact of a simultaneous arrival of the signals at the middle of the train.

Observer 1. The light signals must travel equal distances until they reach me. Consequently, they are sent simultaneously.

Observer 2. The light signals reached me when the middle of the train was moving past me. Consequently, they were sent somewhat earlier. But "earlier" the head of the train was closer to me than the tail. Therefore, the signal from the tail had to be sent slightly in advance in order to reach me simultaneously with the signal travelling from the head. Consequently, the signal from the tail was sent earlier than that from the head.

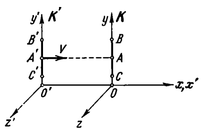


Fig. 2.1. The "imaginary experiment" permitting one to establish that the length of the rulers, oriented at right angles to the direction of the relative motion of coordinate systems, does not vary when measured in any IFR.

It is evident from this plain reasoning that two simultaneous events in one reference frame, that is the train's frame in our example, are far from being simultaneous in another, which in our example is the frame fixed to the Earth.

All subsequent imaginary experiments will be of quantitative nature. In all of them we shall be considering two IFRs designated by K and K' with their relative velocity directed along the common axis x, x' (see Fig. 1.2). It is assumed that the Cartesian axes of the two frames coincide at the initial moment $t = t' = 0$.

(a) *A comparison of the lengths of parallel rulers oriented in the direction perpendicular to that of the relative motion of two IFRs.* Let us take rulers of the same length in each reference frame K and K' and place them along the corresponding axes y and y'. The equal rulers BC and $B'C'$ are illustrated in Fig. 2.1. A and A' are the middle points of the rulers in each reference frame. Let the rulers move so that when the axes y and y' coincide, the middle points A and A' also coincide. The frames K and K' are geometrically identical at the moment $t = t' = 0$. The question is: what quantity will be obtained for the ruler's length $B'C'$ as a result of measurements by the observer from the frame K, and what for the ruler BC, when measured by the observer from the frame K'? The observers have to mark the positions of

the two ends of the rulers moving past them simultaneously in their respective reference frames. For the case considered here synchronism is conveniently established as follows. When the points C' and B' find themselves on the y axis, light signals are sent to the point A'. In the frame K' the sections $A'C'$ and $B'A'$ are equal, the velocity of light c is the same, so that both signals will reach the point A' simultaneously. Consequently, the points C' and B' will cross the y axis simultaneously in the frame K'. Exactly in the same manner the points C and B will cross the y' axis simultaneously in the frame K' as seen from the frame K. Now let us measure the length of the ruler $B'C'$ in terms of the frame K and the length of the ruler BC in terms of the frame K' at the moment $t = t' = 0$, when the y and y' axes coincide. In this case all four points C, C', B, B' find themselves on the common y, y' axis, and the observers in the two frames can compare their results. If it turned out that $CB > C'B'$, or vice versa $C'B' > CB$, it would be possible to detect the difference between the reference frames K and K'. This is inadmissible due to the initial assumption about the equivalence of all inertial frames of reference. That is why the observers from the frames K and K' can only certify that $CB = C'B'$.

Consequently, the lengths (and the length units) oriented in a direction perpendicular to that of the relative motion remain constant when measured in any IFR. But this means that the coordinates of points along the axes perpendicular to the motion direction also remain invariable. Thus, exactly like in the Galilean tranformation

$$y' = y, \quad z' = z. \tag{2.1}$$

(b) *The comparison of clock rates in the frames K and K'.* Observing clock rates in the two frames K and K' moving relative to each other, one can only compare readings of one clock from one frame with readings of several clocks from another frame, because two clocks from different reference frames get together at the same point in space only once. In one of the frames there must be at least two clocks which are supposed to be synchronized in the way described in § 2.2 of this chapter. For the sake of definiteness we shall be comparing one clock from the frame K' with two clocks from the frame K.

Let a clock and a light source be located at the origin O' of the frame K' (Fig. 2.2a). A mirror is set on the z' axis at the distance z_0' from the light source (and the clock) in the direction perpendicular to that of the relative motion. A light signal is transmitted from the source to the mirror from which it is reflected back and returns to the point O' in the time interval $\Delta t' = 2z_0'/c$. Both the

light source and the mirror are at rest in the frame K' and the signal travels "there" and "back" along the same straight line, i.e. the z' axis.

Now let us consider the propagation of the same signal in the frame K relative to which the source and the mirror move to the right together with the frame K' at the velocity V. Although the signal was sent from the two coincident origins O and O', the reflection from the mirror will occur at some other point x_1 of the frame K and the reception of the reflected signal at the point x_2 of the x axis. In this way the path of the signal in the frame K

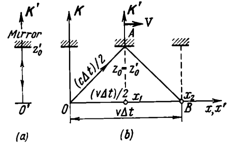


Fig. 2.2. The "imaginary experiment" showing that the interval between two events measured in terms of proper time is always less than the time interval between the same events registered by means of two clocks of any other reference frame. (The "light clock" experiment.) (a) The calculation of the proper-time interval between the sending and reception of a light signal at the origin O' of the coordinate system (b) The calculation of the time interval between the same events in the reference frame K relative to which a light source and a mirror move.

traces out the two sides of an equilateral triangle. As the path travelled by light in the frame K is greater than that in the frame K', one can expect that the time interval Δt between the sending and reception of the signal, when measured in the frame K, will be greater than $\Delta t'$. Indeed, the observer from the frame K will certify that the two events, i.e. the emitting of light from the point O' and its return to the point O', occur at the two different points of space O and B (Fig. 2.2*b*). The time interval Δt between these two events in the frame K will be measured in this case by the two clocks removed from each other by the distance $V\Delta t$ along the motion direction. The velocity of light is equal to c in all reference frames. Therefore, having divided the length of the lateral sides of the triangle OAB by the velocity of light c, we obtain the time interval Δt expressed implicitly:

$$\Delta t = 2\sqrt{z_0^2 + \left(\frac{V\Delta t}{2}\right)^2}/c.$$

Finding Δt from the last equation, we get

$$\Delta t = \frac{2z_0}{c} \frac{1}{\sqrt{1 - \frac{V^2}{c^2}}} = \Gamma \frac{2z_0}{c},$$

where $\Gamma = \left(1 - \frac{V^2}{c^2}\right)^{-1/2}$.

Considering that $z_0 = z_0'$, it follows that

$$\Delta t = \frac{\Delta t'}{\sqrt{1 - \frac{V^2}{c^2}}} = \frac{\Delta t'}{\sqrt{1 - \mathrm{B}^2}} = \Gamma\, \Delta t', \tag{2.2}$$

where the designation $\mathrm{B} = V/c$ is adopted. Since both events occurred at the same point in the frame K', they were registered by means of the same clock. A time interval between events registered by means of the same clock (which implies that the events occurred at the same point of space) is referred to as a *proper-time interval* between these events. Of course, a time interval the initial and the final moments of which are registered at different points of the reference frame and, consequently, by means of different clocks will not be a proper-time interval between events. In the example just examined the proper-time interval is equal to $\Delta t'$. It is seen from Eq. (2.2) that a time interval between events is the least when it is determined in such a reference frame where these events happen at the same point in space. As we shall see in § 3.4 it is possible to indicate the conditions providing the existence of a reference frame in which two given events occur at one point.

Thus we have drawn the most important conclusion: a time interval between two events is a relative quantity; its value depends on the choice of a reference frame. Nothing of the sort was ever known in classical physics, where time intervals possessed absolute properties.

This example effectively illustrates that time readings themselves must be different in different frames. When the origins O and O' coincided, the clocks from the frames K and K' located at this point registered, according to our condition, the moments $t_1 = 0$ and $t_1' = 0$. When the light signal returned to O', the clock from the frame K' registered the time $t_2' = t_1' + \Delta t'$. But at the same moment and at the same point there is the clock from the frame K. This clock is not the one located at O, but another one synchronized with it. Its reading will be $t_2 = t_1 + \Delta t$. As we have already established $\Delta t \neq \Delta t'$, i.e. the clock readings are different. This just means that times of events are registered differently in different reference frames. Note that this calculation of clock readings in the frame K corresponds completely to the

rule of clock synchronization proposed by Einstein and described in § 2.2.

(c) *A comparison of the lengths of rulers arranged parallel to the relative velocity direction.* A proper frame of reference with respect to a given object is such a frame in which this object is at rest. It is customary to designate such a frame by K^0. Let us suppose now that a ruler positioned along the x^0 axis is at rest in this frame. Let us designate the length of the ruler in this frame, i.e. the proper length of the ruler, by l_0. To find the length of the ruler in any reference frame, one has to determine the coordinates of the ends of the ruler simultaneously in this frame. One

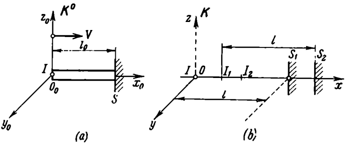


Fig. 2.3. The "imaginary experiment" which permits detecting the "contraction" of the ruler's length when measured in a reference frame in which the ruler moves uniformly and rectilinearly The ruler is oriented in parallel with its motion velocity. (*a*) The measurement of the length of a ruler at rest in the frame K_0 (the proper length of a ruler). (*b*) The measurement of the ruler's length in the reference frame relative to which the ruler moves at the velocity V.

does not care about the simultaneity of measurements only in the frame K^0 in which the ruler rests. Since in everyday life we measure the proper length of objects, the procedure of length measurement is simple and can be performed by means of a direct scale transposition.

A ruler is at rest only in one unique reference frame. In all other inertial frames of reference moving relative to one another the ruler moves, and the direct transposition of a unit scale becomes impossible. Let us resort to the method of length measurement which is also suitable for measuring the length of a ruler moving relative to a reference frame.

Let us place the ruler's left end at the origin O_0 where the light source I is also located. At the ruler's right end the mirror S is fixed perpendicular to the x_0 and x axes (Fig. 2.3*a*). Now let us consider the following two events. The first event: a light signal is sent from the source I toward the mirror along the x_0 axis at the moment $t = t_0 = 0$. The second event: having reflected from the mirror S, the light signal gets back to the ruler's left

end at the point O_0. Both events are registered at the point O_0 by means of one clock. So the time interval between the events is the proper-time interval Δt_0 which can obviously be written down as

$$\Delta t_0 = \frac{2l_0}{c}. \tag{2.3}$$

The same two events look somewhat different to the observer from the frame K (Fig. 2.3*b*). At the moment when the signal is emitted the source I in the frame K is positioned at the point O and the mirror S at the point S_1. By the moment of reflection the mirror will be shifted to the point S_2 and the source I to the point I_1. By the moment of the arrival of the signal reflected from the mirror at the ruler's left end the source will already be shifted to the point I_2. The moments of time corresponding to the first and second events are registered in the frame K at different points and, consequently, by means of different clocks. This means that the time interval Δt between these events can be expressed in terms of Δt_0 according to Eq. (2.2). When light propagates to the right, the velocity at which it overtakes the mirror S is equal to $c - V$, according to the classical velocity summation in the given IFR. When light propagates to the left, it moves toward the mirror at the velocity $c + V$. Designating the ruler's length, unknown so far, by l in the frame K, we obtain the time in which light gets from the source to the mirror, $t_1 = l/(c - V)$, and the time in which light gets from the mirror to the source, $t_2 = l/(c + V)$. Therefore, the time interval between the sending and reception of the light signal in the frame K is

$$\Delta t = t_1 + t_2 = \frac{l}{c - V} + \frac{l}{c + V} = \frac{2l}{c}\frac{1}{1 - \mathrm{B}^2}, \quad \mathrm{B} = \frac{V}{c}.$$

Recalling that $\Delta t_0 = \sqrt{1 - \mathrm{B}^2}\,\Delta t$, and taking into account Eq. (2.3), we obtain from the last equation

$$l = \frac{c\,(1 - \mathrm{B}^2)}{2}\,\Delta t = \frac{c \cdot \Delta t_0}{2}\sqrt{1 - \mathrm{B}^2} = l_0\sqrt{1 - \mathrm{B}^2} = \frac{l_0}{\Gamma}. \tag{2.4}$$

Eq. (2.4) gives the length of the ruler when measured in any inertial frame of reference. In the frame in which the ruler is at rest ($\mathrm{B} = 0$) its length is equal to l_0. It is just the fact from which we started our reasoning. Eq. (2.4) is asymmetric with respect to the lengths l and l_0, since it relates the proper length l_0 of the ruler in the frame K^0 to the improper length l in any other reference frame K.

Thus we have ascertained the relativity of time intervals between events and the relativity of the ruler's lengths or scales directly from Einstein's postulates, the quantities, which, in classical mechanics, were equal in all inertial frames of reference.

These results are inherent in the theory of relativity and require a comprehensive discussion. However, we shall postpone the discussion of the results obtained till §§ 3.2, 3.3, since these results, because of their significance, will be again derived by several methods to reveal some new circumstances essential to the interpretation of Eqs. (2.2) and (2.4).

§ 2.4. The relativity of synchronization of clocks belonging to two inertial frames of reference. The direct derivation of the Lorentz transformation. Up to now we have been considering synchronization of a set of clocks belonging to a given inertial frame. But all inertial frames are equivalent and any event can be registered by an observer located in any inertial frame. An observer marks the coordinates of the event in his own coordinate grid. The set of clocks in every inertial frame of reference registers the time of the event by means of the clock located at the moment of the event at the point in space where this event occurs.

Speaking figuratively, all space is filled up with moving clocks belonging to different reference frames and a momentary light flash at a given point in space, illuminating the dials of all clocks located at it, makes it possible to determine the time of the event, i.e. the light flash, in all reference frames whose clocks were at that point at the moment of the light flash. In order to make out what happens here, it is sufficient to consider the two frames K and K'.

The question is what the clocks of the two frames K and K' will show when they find themselves at one point. Of course, if we want to compare the readings of the clocks from different frames, a certain relationship should be established between the readings of corresponding sets of clocks. The comparison is meaningless without such a relationship. It should be recalled here that all the clocks belonging to each of the frames are synchronized.

It turns out that all one should do is to synchronize only the two clocks, one of the frame K and the other of K', which get together at a given moment. Having synchronized this pair of clocks, we thereby reset the readings of the remaining clocks in each frame. Then it turns out that the clocks of the frames K and K' show different time at all other points in space. This is a very significant result: the clock synchronized in one reference frame is dis-synchronized in terms of any other inertial frame of reference. In other words, if the readings of all the clocks belonging to the frame K are simultaneously fixed in the frame K', it will be seen that all these clocks show different time in the frame K. Now we shall derive the requisite equations.

Usually a relationship between the sets of synchronized clocks in the frames K and K' is established as follows. When the origins of the frames K and K' coincide, the clocks of K and K' located

at the common origin are set to the marks $t = 0$ and $t' = 0$. As we shall soon see, it does not follow from this that the clocks from the frames K and K' show the same time at all other points in space.

We shall need a formula for coordinate transformation for points in space on transition from the frame K to the frame K'. When the origins of coordinates coincide, the coordinate grid of the frame K' is contracted $1/\Gamma$ times in terms of the frame K. The proper unit scales are assumed to be the same in the frames K and K'. Consequently, at the initial moment the coordinates x and x' are related by the ratio (see Eq. (2.4))

$$x = \frac{x'}{\Gamma} \quad \left(\Gamma = \frac{1}{\sqrt{1 - B^2}}\right).$$

By the moment t the coordinate grid of the frame K' will shift as a whole by the distance Vt, so that we obtain $x = x'/\Gamma + Vt$ at that moment. Therefore, if the coordinate of the point in the frame K is equal to x at the moment t, its coordinate x' in the frame K' will be equal to

$$\boxed{x'(x, t) = \Gamma (x - Vt).} \tag{2.5}$$

The coordinate grid does not vary along the y and z axes (§ 2.3), and because of this

$$y' = y, \quad z' = z.$$

Now we are interested in the reading of the clock of the frame K' located at the point x at the moment t. Let us designate this reading by $t'(x, t)$. This quantity can be defined by many methods, but now we shall obtain it using a clock synchronization procedure.

We shall do the following: when the origins of coordinates O and O' coincide and the readings of the two clocks, one from K and the other from K', are equal to zero, a light signal is sent along the common x, x' axis in the direction of the growing values of x, x'. Next, we consider the moment t by the clock of the frame K. At this moment the signal arrives at the point $x_2 = ct$ of the frame K. The arrival of the signal at the point x_2 at the moment t represents the event with the coordinates (x_2, t) in the frame K. In the frame K' the same event will have the coordinates (x_2', t_2'), with $x_2' = ct_2'$ according to the second postulate. But Eq. (2.5) is valid for all events, so substituting x_2 and x_2' into its left-hand and right-hand sides and cancelling by c, we get

$$\boxed{t_2' = \Gamma t (1 - B).} \tag{2.6}$$

This means that a clock belonging to the set of clocks of the frame K' shows at the point x_2 the time t_2' which does not at all

coincide with the time t showed at the same point by a clock of the frame K. Thus we have different time readings; this fact has already been discussed in § 2.3.

Now we can find the reading of still another clock of the frame K' at the moment t. The origin O' will get at the point $x_1 = Vt$ at the moment t. The reference clock of the frame K' will also shift to this point together with the origin. During its shifting it will register the proper-time interval $\Delta t' = t'_1 - 0 = t'_1$.

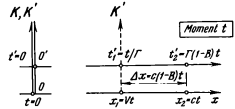


Fig. 2.4. The dis-synchronization of the clocks of the frame K' in terms of the frame K When the origins O and O' coincide, two clocks of the frames K and K', which happen to be at this point, are set so that their readings are $t = 0$ and $t' = 0$ At the moment t (by the frame K clock) the readings of the frame K' clock can be found at the points $x_1 = Vt$ and $x_2 = ct$.

Meanwhile the time interval between the moment when the origins O and O' coincide, and the moment when the origin O' shifts to the point x_1, is equal to $\Delta t = t - 0 = t$, when measured by the clock of the frame K. According to Eq. (2.2)

$$t'_1 = t/\Gamma. \tag{2.7}$$

Thus we have come to the conclusion (Fig. 2.4) that at the moment t (by the clock of the frame K, i.e., simultaneously in the frame K) the clock of the frame K' shows different time when located at different points in the frame K:

$$\text{at the point } x_2 = ct \qquad t'_2 = \Gamma(1 - B)t,$$
$$\text{at the point } x_1 = Vt \qquad t'_1 = t/\Gamma.$$

And all this in spite of the fact that all the clocks from the set of the frame K' have been synchronized within their own frame. But the calculation shows that all these clocks are dis-synchronized in the frame K. We have also found that a dis-synchronization depends on what point of the frame K is selected for clock comparison. Let us find the difference of readings of clocks of the frame K' at the points x_2 and x_1:

$$\Delta t' = t'_2 - t'_1 = \Gamma B t (B - 1).$$

This difference of readings accumulates at the distance $\Delta x = x_2 - x_1 = ct(1 - B)$. Having assumed that a dis-synchronization depends continuously on the distance along the x axis, one can determine a dis-synchronization per unit length:

$$\boxed{\frac{\Delta t'}{\Delta x} = -\Gamma \frac{B}{c}.} \tag{2.8}$$

It is seen from Eq. (2.8) that a dis-synchronization per unit length does not depend on the choice of a moment t, but is determined only by the distance between the clocks of the frame K' as measured in the frame K.

Now one can get for an arbitrary pair of points

$$t_2' - t_1' = -\Gamma \frac{B}{c}(x_2 - x_1).$$

We have already mentioned that sets of synchronized clocks of the frames K and K' are adjusted to each other by setting clocks to the zero reading at the point $x_1 = 0$ at the moment when the coordinate systems of the frames coincide. In other words, the values $t = 0$ and $t' = 0$ are ascribed to the readings of the clocks located at that point. Substituting $x_2 = x$, we obtain from the last equation

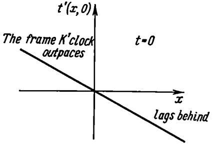


Fig. 2.5. The frame K' clock readings at the moment $t = 0$ (by the frame K clock) at the points of a coordinate x.

$$\boxed{t'(x, t=0) = -\Gamma \frac{B}{c} x.} \tag{2.9}$$

It is seen from Eq. (2.9) what the clock of the frame K' located at the point x will show at the moment $t = 0$ (by the clock of the frame K). Its readings are illustrated graphically in Fig. 2.5. A clock of the frame K' outpaces a clock of the frame K, when located to the left of the origin, and lags behind it to the right of the origin.

Now it is not difficult to ascertain what a clock of the frame K' will show at the moment t, when located at the point x. We shall take advantage of the fact that the difference of the readings of two clocks of the frame K' does not depend on the choice of the moment t. The clock of the frame K' which was at the point $x - Vt$ at the moment $t = 0$, and according to Eq. (2.9) lagged behind the reference clock by the period $\Gamma \frac{B}{c}(x - Vt)$, will get at the point x at the moment t. This clock will always lag behind the reference one by this period. But at the moment t the re-

ference clock will show the time $t_1' = t/\Gamma$ (see Eq. (2.4)), while the clock located at the point x will show the time

$$t'(x, t) = \frac{t}{\Gamma} - \Gamma \frac{B}{c}(x - Vt) = \Gamma\left(t - \frac{B}{c}x\right). \tag{2.10}$$

Eqs. (2.5), (2.6) and (2.10) constitute the Lorentz transformation. Of course, the derivation of Eq. (2.10) may seem clumsy and even superfluous. Indeed, applying the reasoning that has led us to Eq. (2.5) to the transition from the frame K' to K, we obtain

$$x(x', t') = \Gamma(x' + Vt'). \tag{2.11}$$

Solving Eq. (2.11) with respect to t' and substituting x' according to Eq. (2.5), we immediately get Eq. (2.10):

$$t' = \frac{1}{V}\left\{\frac{x}{\Gamma} - x'\right\} = \frac{1}{V}\left\{\frac{x}{\Gamma} - \Gamma(x - Vt)\right\} = \Gamma\left(t - \frac{B}{c}x\right).$$

Having shown the relativity of clock synchronization, we cleared up the physical meaning of different readings of clocks in different inertial frames. Besides, the understanding of "clock dis-synchronization" permits many baffling questions to be avoided. In conclusion, note that the point at which the readings of the clocks of the frames K and K' coincide, moves continuously along the positive x axis at the velocity to be found from Eq. (2.10) by substituting $t = t'$ in it:

$$\frac{x}{t} = \frac{\Gamma - 1}{\Gamma B} c. \tag{2.12}$$

§ 2.5. The Lorentz transformation as a consequence of Einstein's postulates. Let us derive the equations defining the coordinates (x', y', z', t') of an event in the frame K' from the coordinates (x, y, z, t) of the same event in the frame K. These equations performing the transformation of the coordinates of the event must comply with Einstein's postulates.

From the fact that space and time must be uniform in all reference frames it follows that the relationship between the coordinates of the event in two inertial frames must be linear. Indeed, let the origin of the coordinates and the time be changed, i.e. the transformation $x = \bar{x} + x_0$, $y = \bar{y} + y_0$, $z = \bar{z} + z_0$, $t = \bar{t} + t_0$ be performed. If the relationship between the coordinates of the event is linear in the frames K and K', we obtain for x', for example

$$x' = a_1 x + a_2 y + a_3 z + a_4 t =$$
$$= a_1\bar{x} + a_2\bar{y} + a_3\bar{z} + a_4\bar{t} + (a_1x_0 + a_2y_0 + a_3z_0 + a_4t_0),$$

where a_1, a_2, a_3 and a_4 are constants. From the last equation it is seen that the origin has shifted in the frame K' as well, since the expression in parentheses is the same for all points of the

frame K'. However, such a shifting is immaterial due to the uniformity of space and time in all IFRs.

Now let us introduce into the transformation equation at least one second-order term:

$$x' = b_1x^2 + \ldots = b_1\bar{x}^2 + 2b_1\bar{x}x_0 + \ldots$$

Then the second term in the third link of this equaton will depend on $\bar{x}$ and give rise to a distortion or deformation of space. This cannot be tolerated. Consequently, the transformation to be derived is linear.

We shall make use of the arrangement of reference frames illustrated in Fig. 1.2. The relative velocity of the frames is directed along the common x, x' axis with the y and z axes parallel to the y' and z' axes respectively. At the moment $t = t' = 0$ the coordinate systems coincide. The velocity of the frame K' relative to the frame K is equal to V. The x axis results from the intersection of the planes $y = 0$ and $z = 0$. So if the x and x' axes coincide, then $y' = 0$ and $z' = 0$ due to the condition $y = 0$ and $z = 0$. Thus, the transformation equations for the variables y and z must have the following form:

$$y' = Az + By, \quad z' = Cy + Dz.$$

Here A, B, C and D are constants. Since spatial rotations of coordinate systems are inessential for a description of physical phenomena, one can ensure, through rotation of the y' and z' axes around the x' axis, that the plane $y = 0$ is transformed into the plane $y' = 0$ and the plane $z = 0$ into the plane $z' = 0$. Thus, one can put $y' = By$ and $z' = Dz$. However, since the directions y and z are equivalent, i.e. space is isotropic, and the relative velocity of the reference frames is directed along the x, x' axis, it should be true that $B = D$. Hence,

$$y' = Dy, \quad z' = Dz.$$

It remains to determine the coefficient B. Let us consider a unit ruler located in the frame K along the y axis with the coordinates of its ends being $y_1 = 0$, $y_2 = 1$. These coordinates would be $y_1' = 0$, $y_2' = D$ in the frame K', and its length would be $l' = y_2' - y_1' = D$. If we took a unit ruler located in the frame K' along the y' axis $(y_1' = 0,\ y_2' = 1)$, the coordinates of its ends in the frame K would be $y_1 = 0$, $y_2 = 1/D$ and the length would be $l = y_2 - y_1 = 1/D$.

Thus, when measuring a unit ruler of the frame K, an observer from the frame K' will find its length equal to D, while an observer from the frame K measuring a unit ruler in the frame K' will find its length equal to $1/D$. Since all inertial frames are

equivalent, such a result cannot be tolerated unless $D = 1$. Therefore,

$$y = y', \quad z = z',$$

just as it was directly obtained from Einstein's postulates (see § 2.3).

Now let us derive the transformation equations for the variables x and t. Since the transformation is linear,

$$x' = A_{11}x + A_{12}t + A_{10}, \tag{*}$$

and vice versa

$$x = A_{21}x' + A_{22}t' + A_{20},$$

where all coefficients A are constants. From the initial conditions $t = 0$ and $t' = 0$ when the origins O and O' coincide. Hence, $A_{10} = 0$ and $A_{20} = 0$. When observing the point O' we can say that its coordinate x is equal to Vt at the moment t. So from Eq. (*) we get

$$0 = A_{11}Vt + A_{12}t;$$

consequently, $A_{12}/A_{11} = -V$. Designating* A_{12} by Γ', we can rewrite Eq. (*) in the form (see § 2.4):

$$x' = \Gamma'(x - Vt), \tag{2.13}$$

and from the analogous reasonings

$$x = \Gamma(x' + Vt'). \tag{2.14}$$

Thus, the problem has reduced to the determination of the coefficients Γ and Γ'. Due to the uniformity of time and space and to the isotropy of space, both these coefficients can only depend on the absolute value of the velocity V.

It is easy to see that $\Gamma = \Gamma'$. Indeed, let the scale located along the x axis in the frame K have the proper length l_0. If one of its ends is placed at the origin of the frame K, the coordinates of its ends will be $x_1 = 0$ and $x_2 = l_0$ respectively. According to Eq. (2.14) $x_1' = 0$, $x_2' = l_0/\Gamma$ at the moment $t = t' = 0$. (Recall that the two frames coincide geometrically at that moment.) Consequently, the length of the scale in terms of the frame K' is equal to $l' = x_2' - x_1' = l_0/\Gamma'$. Let us take the scale of the same length fixed to the frame K' and also located along the x axis. Then the coordinates of its ends will be $x_1' = 0$ and $x_2' = l_0$. But in terms of the frame K these coordinates will be $x_1 = 0$ and $x_2 = l_0/\Gamma'$ at the moment $t = 0$ according to Eq. (2.13). Consequently, its

* It will be shown very soon that the quantities Γ and Γ' introduced here are equal and coincide with the quantities Γ used in Eqs. (2.2) and (2.4).

length is equal to $l = x_2 - x_1 = l_0/\Gamma'$, i.e. the scale is contracted Γ' times. Since the frames K and K' are equivalent and their relative velocity is the same, the contraction must be identical and, consequently, $\Gamma = \Gamma'$.

Now let us define the quantity Γ. The difference between the frames K and K' lies only in their relative motion, and Γ can depend on the absolute value of the velocity V alone. Let us take advantage of the postulate on the invariant velocity of light *in vacuo* in all IFRs. Suppose, at the moment $t = t' = 0$, when the origins O and O' of the two frames coincide, a light signal is sent from the common origin. Let the event consist in the arrival of the signal at some moment (t in the frame K or t' in the frame K') at some point (x in the frame K or x' in the frame K') located at the x axis. In the frame K this point has the coordinate $x = ct$, whereas in the frame K' the same point has the coordinate $x' = ct'$. These times and coordinates are interrelated by Eqs. (2.13) and (2.14), so that substituting these expressions for x and x' in Eqs. (2.13) and (2.14), we obtain

$$ct' = \Gamma t (c - V), \quad ct = \Gamma t' (c + V).$$

Multiplying term-by-term the left-hand and right-hand sides of these two expressions and cancelling them by tt', we get

$$\Gamma = \frac{1}{\sqrt{1 - \frac{V^2}{c^2}}} = \frac{1}{\sqrt{1 - B^2}}, \quad B = \frac{V}{c}. \tag{2.15}$$

We have thus seen that this quantity Γ coincides with the quantity Γ which first appeared in Eqs. (2.2) and (2.4).

In order to find a transformation equation for the variable t we shall define t' from Eq. (2.14), taking into account Eq. (2.13):

$$t' = \frac{x}{\Gamma V} - \frac{x'}{V} = \frac{x}{\Gamma V} - \frac{\Gamma (x - Vt)}{V} = \Gamma \left\{ t + \frac{x}{V} \left(\frac{1}{\Gamma^2} - 1 \right) \right\} = \Gamma \left(t - \frac{V}{c^2} x \right).$$

Thus, we finally obtain the transformation equations in the following form:

$$x' = \Gamma (x - Vt), \quad y' = y, \quad z' = z, \quad t' = \Gamma \left(t - \frac{B}{c} x \right). \tag{2.16}$$

Eq. (2.16) is referred to as the Lorentz transformation. It is easy to rewrite it for an arbitrary direction of the relative velocity V of the frames K and K'. Indeed, we know that coordinates change in a motion direction and remain invariable in a direction perpendicular to motion. Let us resolve the radius vector $\boldsymbol{r}$ of the

point into components: one parallel to the motion direction $r_{\parallel}$ ($r_{\parallel} \parallel V$) and another perpendicular to it $r_{\perp}$ ($r_{\perp} \perp V$):

$$r = r_{\perp} + r_{\parallel}.$$

Then

$$r'_{\parallel} = \Gamma (r_{\parallel} - Vt), \quad r'_{\perp} = r_{\perp}, \quad t' = \Gamma\left(t - \frac{r_{\parallel}V}{c^2}\right)$$

but

$$r_{\parallel} = V\frac{rV}{V^2}, \quad rV = r_{\parallel}V.$$

Therefore

$$r'_{\perp} = r'_{\parallel} + r'_{\perp} = \Gamma (r_{\parallel} - Vt) + r_{\perp} = \Gamma [(r_{\parallel} + r_{\perp}) - Vt] - - (1 - \Gamma) r_{\perp},$$

$$r_{\perp} = r - r_{\parallel} = r - V\frac{rV}{V^2} = \frac{[V(rV)]}{V^2}$$

and, consequently,

$$r' = \Gamma (r - Vt) + (\Gamma - 1)\frac{[[rV]V]}{V^2},$$

$$t' = \Gamma\left(t - \frac{rV}{c^2}\right).$$

This is the Lorentz transformation in a vector form for an arbitrary direction of the relative velocity. The equation for r' corresponds to the classical Eq. (1.1) and transforms into it when $\Gamma = 1$.

Once again we shall postpone discussing the meaning of the Lorentz transformation (Eq. (2.16)) till we derive it by still another method. That method will lead us to a realization that the real physical world in which all phenomena of nature occur is a four-dimensional manifold, the so-called space-time. The special theory of relativity will appear before us as the theory of four-dimensional space-time, as well as the theory possessing an obvious geometrical meaning. Due to its physical scope and a possibility of a further generalization such an approach proved to be of extreme importance to our whole vision of the world and the first step toward the creation of the theory of gravitation.

§ 2.6. The propagation of the light wave profile. An interval between events. Let us conduct another imaginary experiment, considering it in terms of two IFRs, K and K', the frame K' moving along the common x, x' axis at the velocity V *in vacuo*. At the initial moment $t = t' = 0$, when the origins O and O' coincide, a light flash is triggered. According to the second Einstein postulate light propagates in all directions in the frames K and K' at the same velocity c. Consequently, the wave profile, i.e.

the surface of equal phases, will look like a sphere in each of the frames K and K'. The equation of this sphere can be easily written down:

In the frame K	In the frame K'
$x^2+y^2+z^2=c^2t^2.$	$x'^2+y'^2+z'^2=c^2t'^2.$

Even if we forget everything that was spoken about different time readings t and t' in the frames K and K', we still can explain now why we wrote t' instead of t for the frame K'. Let us suppose the time in the frames be equal, i.e. $t=t'$. Then the radii of the spheres turn out to be equal at a given moment t. Thus, the same physical object, the wave profile, is equally described by the two spheres of equal radii with their centres located at the two points O and O'. This is an absurdity. Hence, one cannot assume $t=t'$. Let us put down the equations in the form

$$c^2t^2-(x^2+y^2+z^2)=0,$$
$$c^2t'^2-(x'^2+y'^2+z'^2)=0.$$

In this imaginary experiment we deal, in fact, with two events. The first one consists in sending a signal from the origin $x_0=0$, $y_0=0$, $z_0=0$ at the moment $t_0=0$, and the second in the arrival of the signal at an arbitrary point of the sphere having the coordinates x, y, z at the moment t. If one makes up the expression

$$\sqrt{c^2(t-t_0)^2-(x-x_0)^2-(y-y_0)^2-(z-z_0)^2}=$$
$$=\sqrt{c^2t^2-x^2-y^2-z^2},$$

that is referred to as the *interval* between these two events and designated by s, the result obtained can be formulated as follows: the square of the interval between the two events, consisting in sending a signal from one point and its arrival at another, must be equal to zero in any reference frame:

$$s^2=0,\quad s'^2=0. \tag{2.17}$$

Of course, the interval between the events can be found not only for the sending and arrival of a light beam. If the coordinates of Event 1 are defined by the numbers x_1, y_1, z_1, t_1 and the coordinates of Event 2 by the numbers x_2, y_2, z_2, t_2, the interval s_{12} between these events is equal to

$$s_{12}=\sqrt{c^2(t_2-t_1)^2-(x_2-x_1)^2-(y_2-y_1)^2-(z_2-z_1)^2}.$$

The interval s_{12} for arbitrary events, however, is not equal to zero.

Frequently it is convenient to consider events occurring at infinitely near points and at infinitely near moments. Assuming

in this case $t_2 - t_1 = dt$, $x_2 - x_1 = dx$, $y_2 - y_1 = dy$, $z_2 - z_1 = dz$, we obtain the interval squared in the form

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2.$$

As we shall illustrate now, the basic property of the interval between events is its invariance on transition from one inertial frame to another.

According to Eq. (2.17) it follows from this imaginary experiment, that is the sending and reception of a light signal, that if $ds^2 = 0$ in one IFR, $ds'^2 = 0$ in any other. Both ds and ds' are infinitesimal quantities of the same order and consequently must be proportional to each other. Therefore, one may put down

$$ds^2 = a\, ds'^2,$$

where a is a proportionality factor. This relationship must be valid for the interval between any pairs of events. Indeed, there are no conditions imposed on the relationship between the intervals ds and ds' for a pair of arbitrary events. As to the special events, that is the sending and reception of light signals, the relationship has to be just like it is shown above.

The coefficient a cannot depend on the coordinates x, y, z and the time t. Otherwise it would mean that different points in space and different moments of time are not equivalent. Since we regard space and time as uniform, a must be a constant depending only on the absolute value of the relative velocity of the two IFRs under consideration. Indeed, the coefficient a cannot depend on the direction of the relative velocity of the two IFRs, because otherwise it would imply inequivalence of different directions in space. Due to the isotropy of space we have to presume that a can only depend on the absolute value of the relative velocity of the inertial frames of reference in question.

Let us consider three IFRs, having designated them K, K' and K'' respectively, with $\mathbf{V}_1$ being the velocity of the frame K' relative to K and $\mathbf{V}_2$ the velocity of K'' relative to K. We can write that

$$ds^2 = a(V_1)\, ds_1^2, \qquad (*)$$

$$ds^2 = a(V_2)\, ds_2^2. \qquad (**)$$

Considering directly the frames K' and K'', one can write that

$$ds_1^2 = a(V_{12})\, ds_2^2,$$

where V_{12} is the absolute value of the velocity of the frame K' relative to the frame K''. Substituting the last expression in Eq. (*) and comparing it with Eq. (**), we find that

$$\frac{a(V_2)}{a(V_1)} = a(V_{12}). \qquad (***)$$

Since V_{12} depends not only on the absolute values of the vectors V_1 and V_2 but also on the angle between them (which does not enter explicitly in the last equation), this relationship can be evidently satisfied only when the coefficient a is reduced to a constant value. From the last equation it is clear that the constant a can be equal only to unity. Hence,

$$ds^2 = ds'^2;$$

from the equality of infinitesimal intervals it follows that

$$s = s',$$

i.e. the interval a is invariant with respect to the transformation of coordinates and time, complying with the Einstein postulates. (Note that the intervals s and s' cannot differ by an arbitrary constant, since from $s = 0$ it follows that $s' = 0$.) We have already seen and shall make sure again that such a transformation is the Lorentz transformation.

Thus, the expression $c^2t^2 - x^2 - y^2 - z^2$ must remain invariable on transition from the frame K to K'. When the frames K and K' are arranged the way it is shown in Fig. 1.2, then $y = y'$, $z = z'$ and the sum $y^2 + z^2$ becomes an invariant. In this case the expression

$$s^2 = c^2t^2 - x^2 \tag{2.18}$$

will be, in fact, the transformation invariant.

§ 2.7. The Lorentz transformation as a consequence of the invariance of the interval between events. In the previous section it was shown that the coordinates of two events must satisfy the equation

$$c^2t'^2 - x'^2 = c^2t^2 - x^2, \quad \text{or} \quad x'^2 - c^2t'^2 = x^2 - c^2t^2 \tag{2.19}$$

on transition from one inertial frame of reference to another. Here we suppose for the sake of simplicity that one of the events has the coordinates (0, 0, 0, 0), and due to our agreement about reference frames it means that this event also has the coordinates (0, 0, 0, 0) in another frame.

Let us take one more step to simplify the notation. The reader has evidently already noticed how common is the product of the velocity of light by the time ct. Let us introduce a new time unit, a *light metre*, which is the time interval that light requires to propagate over the distance 1 m. Obviously, 1 m of time = 1 m/s, i.e. 1 m is covered in $1/c$ seconds.

Light travels τ metres in the time $\tau \cdot 1/c = t$ s. Hence, it is clear that

$$\tau\,(\text{light} \cdot \text{m}) = ct\ (\text{s}). \tag{2.20}$$

This unit will not appear exceptional if one recalls that in astronomy distances are measured in terms of time (and the velocity of light), that is in light years.

So, if time is measured in light metres, the expression for the invariant interval between events becomes quite simple:

$$x'^2 - \tau'^2 = x^2 - \tau^2. \tag{2.21}$$

The easiest way to find a transformation satisfying Eq. (2.21) is as follows. We know from § 2.5 that a transformation of coordinates and time must be linear. Let us write down such a transformation using indefinite constant coefficients in the form

$$\begin{aligned} x' &= a_1 x + b_1 \tau, \\ \tau' &= a_2 x + b_2 \tau. \end{aligned} \tag{2.22}$$

Substituting Eq. (2.22) into the left-hand side of Eq. (2.21) and grouping the coefficients at x^2, τ^2 and $2x\tau$, we obtain

$$x'^2 - \tau^2 = x^2 (a_1^2 - a_2^2) - \tau^2 (b_2^2 - b_1^2) + 2x\tau (a_1 b_1 - a_2 b_2) \equiv x^2 - \tau^2. \tag{2.23}$$

The last link of this equation is written down according to Eq. (2.21); it must be identically satisfied for any x and τ. This requires, however, the following equations to be complied with:

$$a_1^2 - a_2^2 = 1, \quad b_2^2 - b_1^2 = 1, \quad a_1 b_1 - a_2 b_2 = 0. \tag{2.24}$$

These equations are very easy to satisfy, having assumed the coefficients a and b equal to hyperbolic functions (defined and described in Appendix 1, § 9):

$$a_1 = \cosh \theta_1, \quad a_2 = \sinh \theta_1, \quad b_2 = \cosh \theta_2, \quad b_1 = \sinh \theta_2.$$

In this case the first two equations of (2.24) are satisfied automatically. It follows from the third equation, rewritten as $a_1/a_2 = b_2/b_1$, that $\tanh \theta_1 = \tanh \theta_2$, and in order to satisfy this equation it is sufficient to assume $\theta_1 = \theta_2 = \theta$. Thus, the transformation (2.22) takes the form

$$\begin{aligned} x' &= x \cosh \theta + \tau \sinh \theta, \\ \tau' &= x \sinh \theta + \tau \cosh \theta. \end{aligned} \tag{2.25}$$

The parameter θ can depend only on the relative velocity V. It is referred to as a velocity parameter and plays an important role in the STR (see § 3.5). To define it, one should use the first equation (2.25). Assuming $x' = 0$ for the origin O', we obtain

$$\frac{x}{\tau} = - \tanh \theta. \tag{2.26}$$

However, the left-hand side of the equation, when written in conventional time units, appears as $x/ct = \mathrm{B}$, since x/t at the origin O' is just equal to the velocity of the frame K' relative to K. So, we have found the relationship between the parameter θ and the velocity V:

$$\tanh \theta = -\mathrm{B}. \tag{2.27}$$

One easily finds from this (see Appendix I, § 9)

$$\begin{aligned} \cosh \theta &= \frac{1}{\sqrt{1-\tanh^2 \theta}} = \frac{1}{\sqrt{1-\mathrm{B}^2}} = \Gamma, \\ \sinh \theta &= \tanh \theta \cosh \theta = -\Gamma\mathrm{B}. \end{aligned} \tag{2.28}$$

Finally, we obtain the sought for transformations

$$\begin{aligned} x' &= x\Gamma + \tau(-\Gamma\mathrm{B}) = \Gamma(x - \mathrm{B}\tau), \\ \tau' &= x(-\Gamma\mathrm{B}) + \tau\Gamma = \Gamma(\tau - \mathrm{B}x). \end{aligned} \tag{2.29}$$

For the reverse transformations we obtain

$$\begin{aligned} x &= \Gamma(x' + \mathrm{B}\tau'), \\ \tau &= \Gamma(\tau' + \mathrm{B}x'). \end{aligned} \tag{2.30}$$

(The easiest way is to replace the primed quantities by the unprimed ones and B with —B.)

These perfectly symmetric equations are easily identified with the same Lorentz transformation of Eq. (2.16); it is enough to make the substitution of Eq. (2.20).

Thus, we have again obtained the Lorentz transformation, proceeding from the invariance of the interval and the uniformity of space and time. There is no wonder in this, because the interval invariance is a direct consequence of the Einstein postulates.

§ 2.8. Complex values in the STR. Symmetric designations. Sometimes for the sake of the formal convenience an imaginary time coordinate $\tilde{\tau} = ict = i\tau$ is introduced. This practice is efficient when used in the framework of the special theory of relativity, because it frees us of necessity to introduce and distinguish co- and contravariant coordinates (see Appendix I, § 8). An introduction of such coordinates is inevitable in relativistic electrodynamics, unless the imaginary time is used. It should be pointed out that an introduction of an imaginary time is only a matter of convenience and one can do without it. Therefore, there is nothing mysterious about the appearance of the number i. In the final form all formulae for coordinates and time do not contain the number i. This confirms once more that it plays only an auxiliary role.

So, for the sake of the formal convenience we introduce an imaginary coordinate $\tilde{\tau} = ict$. Then

$$s^2 = c^2t^2 - x^2 = -(x^2 + \tilde{\tau}^2) \quad (i^2 = -1).$$

Here is the derivation of the Lorentz transformation using the imaginary variable $\tilde{\tau}$. Consider the plane of variables (x, τ). In this plane the expression $x^2 + \tau^2$ represents a distance from the origin of coordinates to the point (x, τ). This distance does not vary on rotation of the coordinate system through the angle φ in the plane (x, τ).

Rotation in a conventional (Euclidean) plane through the angle φ is described by the following equations (see Appendix I, § 2):

$$x' = x \cos \varphi + y \sin \varphi, \quad y' = -x \sin \varphi + y \cos \varphi, \tag{2.31}$$

where all the quantities are real.

Let us consider rotation in the plane (x, τ) for the case when one of the coordinates is purely imaginary. We shall assume that Eq. (2.31) retains its appearance in this case as well. As we shall see later, the geometric meaning of the equations with an imaginary variable differs essentially from that of Eqs. (2.31). So let us put down the sought for transformation in the form

$$x' = x \cos \varphi + \tilde{\tau} \sin \varphi, \tag{2.32a}$$

$$\tilde{\tau}' = \tilde{\tau} \cos \varphi - x \sin \varphi. \tag{2.32b}$$

Let us clear up the meaning of the parameter φ. It can be associated only with the velocity V of a relative motion of the frames K and K', because this is the only parameter by which they differ. Take any point in the frame K' ($x' = \text{const}$). It moves relative to K just as the whole frame does, i.e. at the velocity V. For any point rigidly fixed to the frame K' one can write $V = dx/dt$. Assuming x to be a function of $\tilde{\tau}$, differentiate Eq. (2.32a) with respect to $\tilde{\tau}$. We shall obtain $dx \cos \varphi + d\tilde{\tau} \sin \varphi = 0$, whence it follows that

$$\frac{dx}{d\tilde{\tau}} = \frac{1}{ic} \frac{dx}{dt} = -\tan \varphi,$$

and, consequently,

$$\tan \varphi = i\mathrm{B}. \tag{2.33}$$

The tangent proved to be an imaginary quantity. This reminds us once again that there is an imaginary quantity among the variables.

Using conventional formulas of trigonometry one can find from Eq. (2.33)

$$\sin\varphi = \frac{\tan\theta}{\sqrt{1+\tan^2\varphi}} = \frac{iB}{\sqrt{1-B^2}} = iB\Gamma,$$
$$\cos\varphi = \frac{1}{\sqrt{1+\tan^2\varphi}} = \frac{1}{\sqrt{1-B^2}} = \Gamma, \tag{2.34}$$

where the designation used already in Eq. (2.15), $\Gamma = (1 - B^2)^{-1/2}$, is introduced. Substituting the values of $\cos\varphi$ and $\sin\varphi$ in Eq. (2.32), we obtain the sought for transformation for the variables x, τ:

$$x' = \Gamma(x + iB\tau), \quad \tau' = \Gamma(\tau - iBx). \tag{2.35}$$

The equations for the transformation of the coordinates of the event from the frame K to K' must differ from those for the transformation of the coordinates of the same event from the frame K' to K only by the substitution of the primed values for the unprimed ones and vice versa. Besides, the sign of the velocity V should be changed to the opposite. In this way we obtain from Eq. (2.35)

$$x = \Gamma(x' - iB\tau'), \quad \tau = \Gamma(\tau' + iBx'). \tag{2.36}$$

Of course, the same result would be obtained if Eqs. (2.35) were solved directly with respect to x and τ.

In Eqs. (2.35) and (2.36) one can easily change over to the real variables x and t by substituting $\tau = ict$ and $\tau' = ict'$. Then we directly get the transformation formulae (2.16). The direct and inverse transformations of the variables x and t have the form

$$x' = \Gamma(x - Vt) \quad \text{(a)}, \qquad x = \Gamma(x' + Vt') \quad \text{(c)},$$
$$t' = \Gamma\left(t - \frac{B}{c}x\right) \quad \text{(b)}, \qquad t = \Gamma\left(t' + \frac{B}{c}x'\right) \quad \text{(d)}. \tag{2.37}$$

Subsequently a comparison of the Lorentz transformation given in the form of Eqs. (2.25) and (2.32) will be of use to us. Recall that in Eq. (2.25) all the quantities are real, while in Eq. (2.32) a time variable is imaginary. Having made use of the relationships (see Appendix I, § 9)

$$\cos i\alpha = \cosh\alpha, \quad \sin i\alpha = i\sinh\alpha, \quad \tan i\alpha = i\tanh\alpha,$$

we see that it is sufficient to substitute $\varphi = -i\theta$ in Eq. (2.25) in order to convert it into Eq. (2.32). Thus, in a plane of real variables x, t we deal in formal terms with the rotation of a Cartesian system through an imaginary angle. Such a rotation resembles very little a true rotation of a Cartesian system, and Eqs. (2.25) defining it are only a "parody" of Eqs. (2.31) describing a true rotation. We shall explain somewhat later in this section the

geometrical meaning of the "rotation" of the x, τ axes according to Eqs. (2.32). and now we shall derive the Lorentz transformation in a symmetric form to be used hereinafter.

Let us introduce the symmetric designations of the basic variables as follows:

$$x_1 = x, \quad x_2 = y, \quad x_3 = z, \quad x_4 = ict = i\tau \tag{2.38}$$

for imaginary time and

$$x^0 = ct = \tau, \quad x^1 = x, \quad x^2 = y, \quad x^3 = z \tag{2.39}$$

for real time.

The set of variables (Eq. (2.38)) is convenient to use when relativistic electrodynamics is described. As to the set of variables of Eq. (2.39), it is adopted in the book [9] which contains a description of the general theory of relativity. A transition from the STR to the general theory of relativity is more expedient to make without resorting to the number i. Let us rewrite the corresponding transformation of variables (Eq. (2.30)) in the real form:

$$\begin{aligned} x^0 &= \Gamma x^{0'} + \Gamma B x^{1'} + 0 \cdot x^{2'} + 0 \cdot x^{3'}, \\ x^1 &= \Gamma B x^{0'} + \Gamma x^{1'} + 0 \cdot x^{2'} + 0 \cdot x^{3'}, \\ x^2 &= 0 \cdot x^{0'} + 0 \cdot x^{1'} + 1 \cdot x^{2'} + 0 \cdot x^{3'}, \\ x^3 &= 0 \cdot x^{0'} + 0 \cdot x^{1'} + 0 \cdot x^{2'} + 1 \cdot x^{3'}, \end{aligned} \tag{2.30$'$}$$

and using imaginary time (Eq. (2.36)):

$$\begin{aligned} x_1 &= \Gamma x_1' + 0 \cdot x_2' + 0 \cdot x_3' - iB\Gamma x_4', \\ x_2 &= 0 \cdot x_1' + 1 \cdot x_2' + 0 \cdot x_3' + 0 \cdot x_4', \\ x_3 &= 0 \cdot x_1' + 0 \cdot x_2' + 1 \cdot x_3' + 0 \cdot x_4', \\ x_4 &= iB\Gamma \cdot x_1' + 0 \cdot x_2' + 0 \cdot x_3' + \Gamma \cdot x_4'. \end{aligned} \tag{2.36$'$}$$

The transformations (2.30′) and (2.36′) can be written in the abbreviated form:

$$x^i = \alpha_{ik} x^{k'} \quad (a), \quad \Big| \quad x_i = \tilde{\alpha}_{ik} x_k' \quad (b). \tag{2.40}$$

In Eqs. (2.40a) and (2.40b) summation is performed for an index k running from 0 to 3 in (a) and from 1 to 4 in (b). The index i is "free" taking on all values from 0 to 3 in (a) and from 1 to 4 in (b), the coefficients α_{ik} and $\tilde{\alpha}_{ik}$ making up the matrices

$$\alpha_{ik} = \begin{pmatrix} \Gamma & \Gamma B & 0 & 0 \\ \Gamma B & \Gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (a) \quad \Bigg| \quad \tilde{\alpha}_{ik} = \begin{pmatrix} \Gamma & 0 & 0 & -iB\Gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ iB\Gamma & 0 & 0 & \Gamma \end{pmatrix} \quad (b), \tag{2.41}$$

respectively which are referred to as the Lorentz transformation matrices. Matrices of this form are always used for transformation of coordinates and time on transition from one inertial frame of reference to another. These matrices differ only in the value of the relative velocity V, i.e. in different values of $B = V/c$.

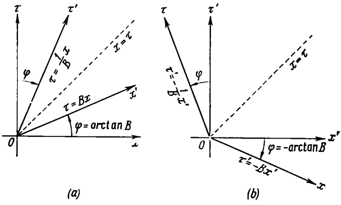


Fig. 2.6. A geometric illustration of the Lorentz transformation. The Lorentz transformation reduces to the rotation of the x and τ axes through the angle $\varphi = \arctan B$ around the origin in the direction of the coordinate angle bisector to their final positions x', τ'. The straight lines $x' = \text{const}$ are now parallel to the $O\tau'$ axis and the straight lines $\tau' = \text{const}$ are parallel to the Ox' axis (we have passed over to the rectilinear oblique-angled system of coordinates). The transition from the frame K to K' corresponds to the convergence of the τ' and x' axes (*a*); the inverse transition to the divergence of the τ and x axes (*b*).

The equations of an inverse transition, i.e. a transition from the frame K to K', are obtained by substituting $-B$ for B. Let us designate the matrix of the transition from the frame K to K' by α'_{ik}, so that

$$x^{i'} = \alpha'_{ik} x^k, \quad x'_i = \tilde{\alpha}'_{ik} x_k. \tag{2.42}$$

For the matrix with the real elements the indicated substitution brings about a completely new matrix α'_{ik}. But the matrix α'_{ik} turns into $\tilde{\alpha}_{ki}$ (with lines and columns transposed), when $-B$ is substituted for B, so that

$$x'_i = \tilde{\alpha}_{ki} x_k. \tag{2.43}$$

§ 2.9. A geometric illustration of the Lorentz transformation. Since in our choice of a relative position of inertial frames of reference the coordinates y and z do not vary, it is sufficient to consider the transformation of the reference frames in the plane (x, τ).

Let the x and τ axes of the frame K be depicted by two mutually perpendicular straight lines (Fig. 2.6). To draw the axes of the

frame K' in this diagram, let us resort to Eq. (2.29) from which the origins of the reference frames K and K' are seen coinciding (when $x = 0$ and $\tau = 0$, $x' = 0$ and $\tau' = 0$ as well). The point $x' = 0$, i.e. the origin of the frame K' moves at the velocity V relative to the frame K. Hence, its motion is illustrated in this diagram by a straight line making an angle φ with the axis τ, the angle φ being defined by the relation $\varphi = \arctan \mathrm{B}$. But the straight line $x' = 0$ is the time axis in the frame K'. Consequently, the Lorentz transformation of the τ axis reduces to an inclination of the τ' axis by an angle φ to the τ axis.

The x' axis is defined by the condition $\tau' = 0$. But from Eq. (2.29) it is seen that this condition is satisfied in the frame K on the line $\tau = \mathrm{B}x$. Of course, the τ' axis could also be found from the condition $x' = 0$, but then we would obtain the same straight line $x = \mathrm{B}\tau$ from (2.29). Thus the equations for the new axes will be written as follows:

$$\text{axis } \tau'\colon\ \tau = \frac{1}{\mathrm{B}}x; \quad \text{axis } x'\colon\ \tau = \mathrm{B}x. \tag{2.44}$$

The angle between the x and x' axes is defined from the relation $\varphi = \arctan \mathrm{B}$. Thus, the Lorentz transformation reduces to a conversion of the rectangular reference frame x, τ to the oblique-angled one x', τ'; the x and τ axes rotate around the origin in the direction of the bisector of the coordinate angle through the same angle $\varphi = \arctan \mathrm{B}$ (see Fig. 2.6*a*). This is what rotation through an imaginary angle means! In formal terms rotation of a rectangular system just considered does not at all resemble rotation of a Cartesian system of coordinates.

Our result shows that we cannot remain within the framework of orthogonal axes x, τ when considering inertial frames of reference and resorting to a geometric illustration of this transformation. Even if the axes of the initial frame are orthogonal, a transition to any frame K' makes it oblique-angled. Fig. 2.6*b* illustrates the transition from the orthogonal frame K' to K according to Eq. (2.30). But an emergence of oblique-angled coordinates makes it necessary to distinguish between co- and contravariant coordinates (see Appendix I, § 8). That is why it is so difficult to bypass these notions in the STR without hiding beyond the number i (see § 2.8).

CHAPTER 3

CONSEQUENCES OF THE LORENTZ TRANSFORMATION. THE CLASSIFICATION OF INTERVALS AND THE PRINCIPLE OF CAUSALITY. THE K CALCULUS

One should not expect to obtain any new consequences of the Lorentz transformation which are not obtainable directly from the Einstein postulates. In the final analysis the Lorentz transformation itself is a consequence of the Einstein postulates. The beginning of this chapter is indeed dedicated to the analysis of the results that have been already obtained in Chapter 2. Surely, they are obtained in a much simpler way from the Lorentz transformation and we shall take advantage of this fact. We shall not even discard an opportunity to show how all these consequences, including the law of velocity transformation, can be obtained even without resorting to a construction of a coordinate system (the K calculus). Naturally, the question may arise as to why the Lorentz transformation is significant, since all the results obtained until now can be derived by other means. The point is that despite the significance of the results obtained, they are not everything we need yet. In order to get convinced in the validity of the principle of relativity, one has to know how the basic equations of physics are transformed on transition from one IFR to another. It is the Lorentz transformation that makes the basis for such transformations.

§ **3.1. On the measurement of lengths and time intervals. The relativity of simultaneity.** The Lorentz transformation makes it possible to compute the coordinates of an event (including a time "coordinate") on transition from one IFR to another. But an event is only an element of a physical phenomenon, and in the final analysis the major task of the STR is the corresponding computation of the physical quantities observed. However, prior to initiating the task it is necessary to dwell on the measuring procedure for basic physical quantities. The most important physical measurements are those of distances (lengths) and time intervals. We are accustomed to measuring lengths of objects or distances between points, when these objects or points are motionless relative to us. If an object is at rest, it is sufficient to transpose a unit scale along the length to be measured the necessary

number of times. This is how we do it in our everyday life, when measuring, say, cloth or the length of a room.

It is also rather easy to measure a time interval between two events occurring at the same point where a clock is located. One has only to register the moment when the first event occurred and the moment of time when the second one did. The difference of the clock's readings will give the time interval between the events. In just this way, for example, a duration of a lecture or a football match is ascertained.

But how to measure the length of an object moving relative to us? Let a train move past us at a great velocity and we want to determine its length. It is far from being simple for one man to do this. He has to note simultaneously the positions of the head of the train (the locomotive) and the tail of the train relative to some certain motionless points on the ground. But as soon as he notes the position of the locomotive and begins to turn his head, the tail of the train will have gone ahead. Consequently, one should take special care to mark simultaneously the positions of the head and the tail of the train.

Having marked the simultaneous positions of the head and the tail of the train on the ground, we can readily measure the distance by conventional means used for measuring motionless objects.

And how to measure a time interval between events occurring at different points in space? Recall how they measure a time which a sprinter takes to run a hundred metre race. The events in this case are represented by the sportsman's start and finish. And there is only one clock! A starter's shot serves as a signal to start the race and to actuate a timer located at the finish. Sound propagates in air at the velocity of 330 m/s, so that the sportsman will start running before an umpire located at the finish actuates his timer. This is not very essential, though, because the velocity of the runner is very small (at best about 36 km/h = 10 m/s, which is small even relative to the velocity of sound). But from this example one may perceive that the determination of a time interval between two events happening at different points in space requires attention.

In § 2.1 we discussed how such a problem is dealt with in the theory of relativity: each inertial frame has its own coordinate system and motionless clocks are located at all its points, wherever needed. These clocks are synchronized within that reference frame, so that equal readings of the clocks correspond to the same moment of time in the frame.

When we turn to comparing events in the two inertial frames K and K', we link the readings of the synchronized set of clocks of the frame K with the readings of the analogous set of the frame

K' by assuming $t = 0$ and $t' = 0$ at the coinciding origins O and O' (see § 2.2). Recall that the Lorentz transformation is just a conversion of the coordinates of the event in the frame K, i.e. the coordinates (x, y, z, t), into the coordinates of the same event in the frame K'. From Eqs. (2.16) or (2.29), or (2.40), we obtain (x', y', z', t') expressed via (x, y, z, t). Of course, all these surprising, from the viewpoint of the "common sense", consequences of the Einstein postulates that we discussed in Chapter 2, can be obtained from the Lorentz transformation. Now we shall be occupied with this.

First, let us derive two convenient equations that we shall need later. Consider the two arbitrary events I (x_1, y_1, z_1, t_1) and II (x_2, y_2, z_2, t_2). The conversion of coordinates and times of these events into the frame K' yields, according to Eq. (2.37a, b),

$$t'_2 = \Gamma\left(t_2 - \frac{\mathrm{B}}{c} x_2\right), \qquad x'_2 = \Gamma(x_2 - Vt_2),$$

$$t'_1 = \Gamma\left(t_1 - \frac{\mathrm{B}}{c} x_1\right), \qquad x'_1 = \Gamma(x_1 - Vt_1).$$

Making up the differences $t'_2 - t'_1$ and $x'_2 - x'_1$, i.e. subtracting the lower equations from the upper ones, and designating $\Delta x = x_2 - x_1$, $\Delta x' = x'_2 - x'_1$, $\Delta t' = t'_2 - t'_1$, $\Delta t = t_2 - t_1$, we obtain the necessary equations (the inverse transition equations are also written out):

$$\Delta x' = \Gamma(\Delta x - V\Delta t), \quad (3.1) \qquad \Delta x = \Gamma(\Delta x' + V\Delta t'), \quad (3.1')$$

$$\Delta t' = \Gamma\left(\Delta t - \frac{\mathrm{B}}{c}\Delta x\right), \quad (3.2) \qquad \Delta t = \Gamma\left(\Delta t' + \frac{\mathrm{B}}{c}\Delta x'\right). \quad (3.2')$$

In fact, Eqs. (3.1) and (3.2) as well as Eqs. (3.1′) and (3.2′) are the Lorentz transformations for differences of spatial coordinates and times of the two events. These equations should be supplemented with the relations $\Delta y' = \Delta y$ and $\Delta z' = \Delta z$.

It immediately follows from Eq. (3.2) that two simultaneous events in the frame K are not simultaneous in the frame K'. Indeed, assuming $\Delta t = 0$ in Eq. (3.2), we obtain

$$\Delta t' = -\Gamma \frac{V}{c^2} \Delta x. \tag{3.3}$$

It is seen that $\Delta t' \neq 0$, if $\Delta x \neq 0$. But if $\Delta x = 0$ at $\Delta t = 0$ as well, the events either coincide or happen in the plane $x = \text{const}$. For such events $\Delta t' = 0$.

Two conditions follow from Eq. (3.2). Provided they are satisfied, one may ignore the relativity of time intervals between events. First, one should suppose $\mathrm{B} \ll 1$; then $\Gamma \approx 1$ and the following expression can be written: $\Delta t' = \Delta t - (V/c)(\Delta x/c)$. The second term in this expression can be ignored if the ratio $\Delta x/c$

is rated small. This second condition is positively satisfied if events occur in a limited region of space along the x axis, i.e. at small Δx. No limitations are imposed, however, on the region of space along the directions y and z, since all events in the plane $x = \text{const}$ happen at the same moment of time according to the clock of the frame K'.

Certainly, the relativity of simultaneity reveals itself in the dis-synchronization of clocks that we examined in detail in § 2.4. It is sufficient to put $t' = 0$ in Eq. (2.37b), and we immediately obtain Eq. (2.9) which did not come easily before. As soon as we assume $\Delta t = 0$ in Eq. (3.2), Eq. (2.8) is obtained. This example shows how much can be hidden in plain, by appearance, "transformations".

In everyday life a violation of simultaneity is not perceptible: the time difference Δt is proportional to B/c, as it is seen from Eq. (3.2), provided that $\Delta t = 0$ is assumed (simultaneity in the frame K). From the same Eq. (3.2) it is seen that if $\frac{B}{c}\Delta x$ has a substantial value, $\Delta t'$ can also assume a substantial value provided that Δx is great.

It is very important to point out that the relativity of simultaneity is dictated by the finiteness of the velocity of light. If a formal passing to the limit $c \to \infty$ is performed (in fact, it means that $B \to 0$), simultaneity becomes absolute. This result corresponds to the case of small relative velocities of reference frames.

It is seen from Eqs. (3.1) and (3.1′) that two events occurring at points of space with the same x coordinate in the frame K, i.e. at the same point of the frame K, will have different x' coordinates in the frame K'. Indeed, from Eq. (3.1) we get $\Delta x' = \Gamma(\Delta x - V\Delta t) = -\Gamma V \Delta t$. But Δt is a proper-time interval, and so $\Gamma \Delta t = \Delta t'$. Hence, $\Delta x' = -V\Delta t'$. The meaning of the last result is evident; it defines the displacement of the point x relative to the frame K' registered in the frame K'.

§ 3.2. Relativity of length of moving rulers (scales). A visible shape of objects moving at relativistic velocities. Let us consider now a measurement of the length of a moving ruler. Let the clocks in the frame K be synchronized and spatial marks made. Suppose that a ruler oriented along the x axis moves relative to the frame K at the velocity V. One can fix the frame K' to this ruler. How to measure the length of the same ruler in the frame K? Obviously, the coordinates of the ruler's ends have to be determined *simultaneously* in the frame K. This requirement of simultaneity leads us to the strange result which we have already discussed in § 2.3: the ruler's length, when measured in a reference frame relative to which the ruler moves, turns out to be less than the length of the same ruler in the reference frame

where it is at rest. So, let the ruler be at rest in the frame K' and the coordinates of its ends be x_1' and x_2'. By definition, its length in the frame K' called, as we have indicated, the proper length is equal to $x_2' - x_1'$. The proper length of the ruler is designated by l_0, i.e. $l_0 = x_2' - x_1'$. Since the ruler is motionless in the frame K', one may not worry about the simultaneity of measurements of the coordinates of its ends: its length can be measured by any conventional means.

In the frame K the coordinates of the ruler's ends will be determined according to the Lorentz transformation (see Eq. (2.37a)):

$$x_2' = \Gamma(x_2 - Vt_2), \quad x_1' = \Gamma(x_1 - Vt_1).$$

Having formed the difference $x_2' - x_1'$, we obtain

$$x_2' - x_1' = \Gamma\{(x_2 - x_1) - V(t_2 - t_1)\}. \tag{3.4}$$

The proper length of the ruler constitutes the left-hand side of Eq. (3.4). In the braces of the right-hand side there are Δx and Δt for the two events, the position of the ruler's left end being x_2 at the moment t_2 and the position of the ruler's right end being x_1 at the moment t_1 (in the frame K).

The quantity Δx will be the ruler's length in the frame K only if the positions of the ruler's ends are registered simultaneously in this frame. Otherwise we can get any value for Δx. Recall the example cited at the beginning of § 3.1 concerning the determination of the length of a moving train: you have just marked the position of the tail of the train and are slowly turning your head toward the locomotive. If you mark the position of the locomotive and then measure the distance between the marks, the distance measured will be *greater* than the proper length of the train. Now proceed in the opposite direction: first mark the position of the locomotive and then turn your head slowly to the tail of the train. It is easy to contrive that if one turns his head slowly enough, the train's length may even get equal to zero. This is just what Eq. (3.4) expresses. Thus, in order to determine the ruler's length unambiguously in the frame K, one has to consider the two simultaneous events in K: the coincidence of the ruler's left end with a certain spatial mark, say x_1, and that of the ruler's right end with another spatial mark, say x_2. This means that $\Delta x = l$ only if $\Delta t = 0$. But then it follows from Eq. (3.4) that $l_0 = \Gamma l$, where l is the ruler's length determined in the frame K. According to custom the last equation is written as follows:

$$l = l_0\sqrt{1 - \frac{V^2}{c^2}} \equiv l_0\sqrt{1 - \mathrm{B}^2}. \tag{3.5}$$

Of course, under the same conditions of the problem one may make use of the inverse transformation equations:

$$x_2 = \Gamma\left(x_2' + Vt_2'\right), \quad x_1 = \Gamma\left(x_1' + Vt_1'\right).$$

Subtracting the right-hand equation from the left-hand one, we obtain

$$\Delta x = \Gamma l_0 + \Gamma V\left(t_2' - t_1'\right) = \Gamma l_0 + \Gamma V\,\Delta t'.$$

But Δx becomes the length l only under the condition $\Delta t = 0$. Having expressed $\Delta t'$ in terms of Δt and Δx from Eq. (2.37b), $\Delta t' = \Gamma\left(\Delta t - \frac{V}{c^2}\Delta x\right)$, and having assumed $\Delta t = 0$ in this equation, we obtain Eq. (3.5) again.

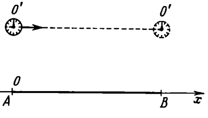


Fig. 3.1. Measurement of the length of a moving ruler.

Here are two more methods of determining the length of a moving ruler which, naturally, bring about the same result. Let the ruler AB be at rest in the frame K, so that its left end A coincides with the origin O (Fig. 3.1). At the moment $t = 0$ the origin O' of the frame K' coincides with O, the clock's reading in O' being $t' = 0$. Then the observer located at B registers the moment when O' passes the point B. Let it be the moment t. The velocity of the frame K' is known. Therefore, the proper length of the ruler $l_0 = V\Delta t$. This is the proper length of the ruler because it is measured by the scale and clocks of the frame K in which the ruler is at rest. The velocity is also determined in the frame K. On the other hand, the observer from the frame K' located at the point O' will register his passing by the points A (at the moment $t' = 0$) and B (at the moment t') using his clock. But the ruler moves past this observer also at the velocity V (in the opposite direction), and he will find the length of the moving ruler to be equal to $l = V\cdot\Delta t'$. But $\Delta t' = (t' - 0)$ is the proper-time interval between the two events, the coincidence of O' with A and then with B. As to $\Delta t = (t - 0) = t$, it is the time interval between the same events in the frame K. According to Eq. (2.2) $\Delta t = \Gamma\,\Delta t'$,

and inasmuch as $\frac{l}{l_0} = \frac{\Delta t'}{\Delta t} = \frac{1}{\Gamma}$,

$$l = \frac{1}{\Gamma} l_0 = l_0 \sqrt{1 - B^2}.$$

The relativity of lengths is a direct consequence of the relativity of simultaneity. Let a ruler be at rest in the frame K', the coordinates of its ends being x_2' and x_1'. The proper length $l_0 = \Delta x' = x_2' - x_1'$.

Let two light bulbs fixed at the ruler's ends in the frame K' flash simultaneously ($\Delta t' = 0$), and let these two events be registered in the frame K. Let us find the distance between the

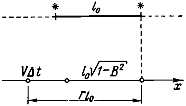


Fig. 3.2. Relativity of the rulers' lengths as a consequence of the relativity of simultaneity of two events.

points at which these events occur in the frame K: $\Delta x = \Gamma(\Delta x' + V\Delta t') = \Gamma\Delta x' = \Gamma l_0$ (see Eq. (3.1')). This means that the distance between the points at which these two events occur is greater than the proper length of the ruler. But this distance is not the ruler's length which will be measured by the observer from K. In order to find the ruler's length in the frame K, one must find the coordinates of the ruler's ends simultaneously in this frame. The events which are simultaneous in the frame K' are delayed relative to one another by $\Delta t = \Gamma\left(\Delta t' + \frac{V}{c^2}\Delta x'\right) = \Gamma\frac{V}{c^2} l_0$ in the frame K (Fig. 3.2). But during this time the ruler's end x_1' will shift in the direction of motion by the distance $V\Delta t = \Gamma B^2 l_0$. Hence the measured length of the ruler will be less than Γl_0 by $V \cdot \Delta t$, i.e.

$$l = \Gamma l_0 - \Gamma B^2 l_0 = l_0 \sqrt{1 - B^2}.$$

We shall also recall that Eq. (3.5) was obtained directly from the Einstein postulates (§ 2.3). Now it is time to dwell on the physical meaning of Eq. (3.5). We have found that the length of a physical object, for example a ruler, is relative, i.e. different in different reference frames. The ruler possesses a maximum length in the frame in which it is motionless, i.e. the proper length

is the greatest. If the ruler's length is determined in an inertial frame, relative to which the ruler moves, its length will prove to be less than the proper length. It follows from Eq. (3.5) that if the ruler were able to move at the velocity c, its length would be equal to zero. But this just cannot happen: any object possessing a finite rest mass, including any feasible frame of reference, cannot reach the velocity c.

What does a contraction of a ruler mean? Frequently one may hear the question as to whether the ruler "actually" gets shorter. To begin with, it is clear that no contraction of the ruler can take place. This follows from the basic principle of the STR, the principle of equivalence of all IFRs. The physical state of the ruler is the same in all IFRs. So there is no question of an emergence of any stresses leading to the ruler's deformation. The ruler's "contraction" comes about solely due to differing methods of length measurements in two reference frames. On the other hand, the observed relativity of the ruler's length is not due to the observer's illusion. This result can be obtained by any reasonable method of measuring the length of a moving object. Moreover, analysing physical phenomena in a given reference frame, the quantity l should be adopted as the object's length according to Eq. (3.5), and not l_0.

It is extremely unfair to speak of the "Lorentz contraction" when alluding to Eq. (3.5), although indeed G. A. Lorentz was the first to suggest this equation in 1892. However, it was interpreted quite differently (see Supplement II) from what we have just discussed.

It was Einstein who clearly said about the reality of the Lorentz contraction: "There is no point to question whether the Lorentz contraction is real or not. The contraction is not real, since it does not exist for an observer moving with the object. However, it is real, since it can be fundamentally proved by physical means for an observer not moving with the object."

Another question is often raised: what is the "actual" length of the ruler? This question has no meaning, if asked "in a broad sense". The question about the ruler's length regardless of a frame does not have any meaning. The ruler has its length in each reference frame; this is just its "actual" length. All inertial frames of reference are equivalent and so are the ruler's length values determined in these frames. In any reference frame the ruler will behave as if it has the length determined in this frame. Although all IFRs are equivalent among themselves, there is still one "selected" coordinate system to which we are accustomed. This is the system in which the ruler is at rest. From the viewpoint of our customary concepts this is just the "actual" length of the ruler. We are prone to adopt it as a true length, but this

length defines the behaviour of the ruler only in this "inherent" reference frame.

Finally, the last remark. The ruler exists objectively, i.e. outside our consciousness and ourselves. But is there any length before measurements are made? A length as a certain number emerges as a result of measurements and a choice of units of length. Of course, the ruler possesses extent, or length, if you want, as a quality before a measurement, but the numerical value of length originates only after measurement. Thus, the numerical value of length of existing objects emerges after measurement, and the result of measurement, as we have established, depends on what kind of instruments are used.

Let us consider rulers having the same proper lengths l_0' and l_0'' in the two reference frames K' and K''. Measurements carried out in the frame K' will give

$$l_0' > l_0''. \tag{3.6}$$

Measurements carried out in the frame K'' will give

$$l_0'' > l_0'. \tag{3.7}$$

The inequalities (3.6) and (3.7) are far from being contradictory because (3.6) was obtained for the scales and clocks of the frame K' and (3.7) for the scales and clocks of the frame K''. The difference in the values of the ruler's length is dictated by the fact that simultaneity in the frames K' and K'' is defined differently. The difficulty in the interpretation of the conclusions of the special theory of relativity lies not in the existence of relative quantities, but in the detection of the equivalence of all inertial frames of reference. The inequalities (3.6) and (3.7) indicate just this equivalence.

The conclusion of the special theory of relativity concerning the relativity of the length of a moving object is unusual partially because in our everyday life we do not perceive such an effect. Let us consider the fastest motion within reach, that is the orbital motion of the Earth. In this case $V = 30$ km/s. The ratio $V/c \approx$ $\approx 10^{-4}$ and $l = l_0\sqrt{1 - 10^{-8}} \approx l_0\left(1 - \frac{1}{2} \cdot 10^{-8}\right) \approx l_0$. It should be stressed again that the contraction of lengths is a direct consequence of the finiteness of the velocity of light. If the velocity of light were infinite, the ruler's length would be the same in all reference frames in accordance with Eq. (3.5). This can also be seen from the fact that in the case of $c \to \infty$ the simultaneity of events becomes absolute.

Although until now we always discussed the relativity of lengths of objects (rulers), it should be borne in mind that actually we dealt with the relativity of distances between two motionless

points in one reference frame when measured by instruments from another frame.

Let us consider a cube at rest in the frame K with the sides Δx, Δy, Δz and the proper volume $\Delta\mathscr{V}_0 = \Delta x\,\Delta y\,\Delta z$. According to the Lorentz transformation in an arbitrary IFR K' we have $\Delta x' = (1/\Gamma)\Delta x$, $\Delta y' = \Delta y$, $\Delta z' = \Delta z$ and, consequently

$$\Delta\mathscr{V}' \equiv \Delta x'\,\Delta y'\,\Delta z' = (1/\Gamma)\,\Delta x\,\Delta y\,\Delta z = (1/\Gamma)\,\Delta\mathscr{V}_0.$$

Hence, the change of the cube's volume on transition from the frame K to the frame K' is determined as

$$\mathscr{V}' = \mathscr{V}_0\sqrt{1 - B^2}. \tag{3.8}$$

It follows from the result obtained that the proper volume of an object is the invariant of the Lorentz transformation:

$$\Gamma'\,d\mathscr{V}' = \Gamma''\,d\mathscr{V}'' = \mathscr{V}_0.$$

Is it possible to observe directly the Lorentz contraction by, say, taking photographs of a rapidly moving object? In the first paper by Einstein dedicated to the theory of relativity one may read the following: a moving body which at rest has a spherical shape is observed from a stationary frame as an ellipsoid with the semi-axes $R(1-\beta^2)^{1/2}$, R, R. The word "observed" here can be interpreted as a visual observation or photographing. For about fifty years after the advent of the STR everybody was sure that a visible shape of a relativistic sphere was an ellipsoid. However, it turned out that the problem of the visible shape of objects moving at a relativistic velocity requires many circumstances to be taken into account, and that a rapidly moving sphere remains spherical. If one assumes that an eye and a photographic plate fix an instantaneous image produced by light, this will mean that the image is produced by rays coming from different sections of the observed object simultaneously on a retina or a photographic plate. But if the optical paths of light going from various points of the observed object are different, a photographic plate will register the positions of the object's points at different moments of time prior to the moment of photographing. The whole effect is caused by the finiteness of the velocity of light. Using a plain example, we shall show why a visible shape of a moving sphere coincides with that of a stationary sphere.

Let us imagine a luminous cube moving along a straight line parallel to one of its faces past a photographic camera or an observer. The photographing or observation takes place at the moment when the centre of the cube crosses the line perpendicular to the motion direction and drawn through the point at which the observer is located (Fig. 3.3a). Naturally, we must

know beforehand that the moving object has the form of a cube in its own reference frame.

At a definite moment all photons emitted simultaneously (in the frame fixed to the plate) at the points of the line AD will reach the plate together with those emitted at the point B earlier by the time interval l/c, l being the edge length of the cube. But at that moment the point B was in the position B'. The simultaneous determination of the positions of the points A and D in the frame fixed to the plate leads, in accordance with the conventional rule of length measurement, to the Lorentz contraction: $l' = l(1-\beta^2)^{1/2}$. On the other hand, $BB' = (l/c)v = \beta l$.

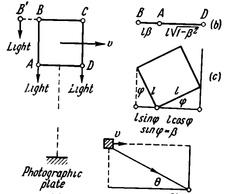


Fig. 3.3. Visual observation of a cube moving past an observer: (a) the mutual disposition of the observer and the cube at $\theta=0$; (b) the visible picture of the moving cube, (c) the possible interpretation of the visible picture by one observer. the rotation of the cube through the angle $\varphi = \arcsin \beta$; (d) the observation of the moving cube at the angle θ

From Fig. 3.3b and c one can realize that the picture of a moving cube that would be seen by a motionless idealized observer coincides with the picture of a motionless cube turned through a certain angle φ. This angle is determined from the relation $\sin\varphi = \beta$. This is a particular case of a more general result: any three-dimensional moving object is seen turned at a given moment. If the cube is so positioned relative to the observer that it is seen at the angle θ' relative to the x' axis when at rest, it will be observed turned through another angle. If the cube is removed far enough from the observer, light travelling from it can be taken for a parallel pencil of rays. When this pencil is observed in the frame K, it propagates at the angle θ to the x axis, as seen by the observer from K, the angles θ and θ' being related by the equation (see Eq. (7.11)):

$$\cos\theta = (\cos\theta' + \beta)/(1 + \beta\cos\theta').$$

A variation of a plane wave front direction on transition from one reference frame to another, moving relative to the former one, is called an aberration of light. The image registered on the plate in this case corresponds to the cube observed in the frame K at the angle θ and turned through the angle $\theta - \theta'$.

From this it is clear that a sphere turns as well, but its outline does not vary, of course (see [28] for details).

And still — can an object be photographed, so that a plate will register a relativistic contraction? To avoid difficulties associated with a turn, one can consider a one-dimensional object which is easy to compare with its own image on the plane. For this purpose the observer in the frame K must know beforehand that a rod moves along a given direction. The rod is at rest in the frame K', and its proper length is also known. In this case the

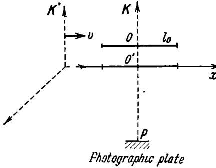


Fig. 3.4. The basic arrangement permitting the Lorentz contraction of a moving rod to be photographed. When the middle of the rod O' gets onto the line PO, a special device opens the shutter at P for an instant to pass through the rays emitted by the rod's points at the moment when the point O' crosses the line PO.

observer constructs a counterpart of the moving rod in his frame K and takes a photograph of the moving rod against the background of its own length.

The simplest arrangement intended for photographing a rod experiencing the Lorentz contraction could be of the type shown in Fig. 3.4. The rod is parallel to the x axis and moves along it. An observer is located on a perpendicular to the x axis, the perpendicular going through the middle of the rod's counterpart at rest in the frame K'. As soon as the middle of the moving luminous rod finds itself on the perpendicular, a mechanism triggers a camera's shutter, and the simultaneous positions of the rod's ends get registered on a photographic plate. As to the stationary counterpart of the rod, it can be photographed at any time, of course. Most likely, this is also an "imaginary experiment".

It should also be mentioned, however, that a "turn of a cube" moving at a relativistic velocity was qualitatively photographed. We refer the reader interested in details to [27].

In general, a "visible" picture may differ quite essentially for different observers. Here is a simple example. Let an observer

be at some distance from a plane at which electric bulbs flash simultaneously in the frame where the plane and the observer are at rest, or the light flash is produced. Then due to the finiteness of the velocity of light the observer will see that the plane starts glowing gradually with an "illumination wave" running from the centre to the periphery. In particular, if an infinite thread is illuminated instantaneously, the remote observer will see two luminous points running apart.

Owing to the same reason a visible (observable) velocity may also differ from a real one: it may prove to be even higher than the velocity of light [33].

§ 3.3. Relativity of time intervals between events. Suppose that two events occurred at some point of the frame K' at time moments t'_1 and t'_2. The time interval separating these events can be registered by the clock located at that point. According to the definition a time interval between events that occurred at the same point of a certain reference frame and registered by the same clock of this frame is called a proper-time interval between the events. Designating the proper-time interval by Δt^0, we obtain in our case

$$\Delta t^0 = \Delta t' = t'_2 - t'_1.$$

Let us define now a time interval between the considered events in the frame K. According to Eq. (3.2) $\Delta t = \Gamma\,\Delta t' = \Gamma\,\Delta t^0$. This result is already familiar to us (Eq. (2.2)).

However, if the events in the frame K' occurred at one point in space, this is not the case in any other frame K. Indeed, let two events occur in the frame K' at different moments but at one point, i.e. $\Delta x' = 0$ but $\Delta t' \neq 0$. According to Eq. (3.1) $\Delta x = \Gamma V$, $\Delta t' \neq 0$. The meaning of the last result is obvious: all points of the frame K' move relative to K at a velocity V, and Δx is just a displacement of any point of the frame K' during the interval considered. Since $\Gamma\Delta t' = \Gamma\Delta t^0 = \Delta t$, then $\Delta x = V\,\Delta t$.

Therefore, time intervals between the same pair of events turn out to be different in different IFRs. We shall observe the least time interval between events in the reference frame in which these events occur at the same point and, consequently, are registered by the same clock. In other words, the proper-time interval is the least.

A time interval between events in terms of the frame K can also be measured as follows. The point of the frame K' at which the two considered events occur, moves at the velocity V relative to the frame K. If the events occurred in the frame K at the points x_1 and x_2, the time interval Δt between the events is, obviously, equal to $\Delta t = \Delta x/V$. But according to Eq. (3.1) $\Delta x =$

$= \Gamma V \Delta t'$ ($\Delta x' = 0$), whence we obtain again

$$\Delta t = \Gamma \Delta t^0 = \frac{\Delta t^0}{\sqrt{1-\beta^2}}. \tag{3.9}$$

The direct experimental confirmation of the conclusion of the STR about the relativity of time intervals is widely known. Light elementary particles (muons) were discovered, on the one hand, in a laboratory as a result of nuclei splitting and, on the other hand, in cosmic rays. The lifetime (the half-life) of muons measured in laboratory conditions proves to be equal to about $2\cdot 10^{-6}$ s. This lifetime can be regarded as a proper lifetime, since the velocities of laboratory muons are non-relativistic (see Eq. (3.9), the velocity of the coordinate system K' is the velocity of the frame fixed to a muon). In the interval $\Delta t^0 = 2\cdot 10^{-6}$ s a muon breaks up into other particles.

It is known that the muons observed in cosmic rays at the surface of the Earth originate in the upper layers of the atmosphere at the height of from five to six kilometres due to the primary cosmic radiation. The velocity of the generated muons moving toward the Earth is comparable with that of light. According to Eq. (3.9) the half-life Δt of a muon in the laboratory frame is equal to $\Delta t = \Gamma \Delta t^0$. In the case of muons $\Gamma \approx 10$ and in the laboratory frame of reference $\Delta t = 2\cdot 10^{-5}$ s. During this time a muon travels the distance $c\Delta t = 3\cdot 10^{10} \times 2\cdot 10^{-5} \approx 6$ km. But for the relativity of time intervals, muons would have travelled only about 600 m and we would not have observed them at the sea level. Thus, only the relativistic transformation of time intervals makes it possible to explain muon showers observed on the Earth.

An excellent illustration of the relativity of time intervals is provided by the Doppler effect. This effect consists in the fact that if a light source and an observer (a receiver) move relative to each other *, the light frequency determined by the observer differs from the frequency that would be observed by him if he was stationary relative to the source. It is natural to refer to a light frequency determined by the observer stationary relative to the source as a proper frequency of light. Let us designate it by ω_0. All our conclusions pertain to vacuum.

Let us first consider the case when the direction of light propagation coincides with the direction of the relative velocity of the source and the observer, the so-called radial Doppler effect.

* Note that in our subsequent discussions of "classical" physics we shall always utilize a typically relativistic assumption that no material medium is needed for light propagation. That is why only the relative velocity of a light source and an observer is essential for us

In Fig. 3.5*a* this corresponds to light propagation along the x, x' axis. To consider the Doppler effect, one can imitate a light wave by sending short pulses from the source at the interval (period) T. Let us suppose now that such pulses are sent from the origin of the frame K, i.e. from the point O. Any observer at rest in this frame will discover that these pulses come to him at the same intervals T. Now let us fix the observer to the frame K'. Let the first pulse be sent from the point O at the moment when the origins O and O' coincide ($t = 0$, $t' = 0$). Naturally, the same pulse will be registered by the observer at the point O'. The next

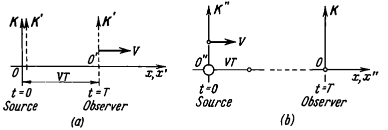


Fig. 3.5. The derivation of equations of the radial Doppler effect: (*a*) an observer moves away from a source; (*b*) an observer approaches a source.

pulse leaves the point O after the time interval T (by the clock of the frame K). But at that moment the origin O' is already at the distance VT from O. The velocity of light relative to O' in the frame K is equal to $c - V$, so that a light ray will need the additional time $VT/(c - V)$ to reach the origin O'. Therefore, the observer at O will receive the pulse after the time interval

$$T' = T + \frac{VT}{c - V} = \left(1 + \frac{V/c}{1 - V/c}\right) T = \frac{1}{1 - B} T \qquad (3.10)$$

when registered by the clock of the frame K. We have obtained the time interval between the first and the second signals received by the observer at O' and registered by the clock of the frame K. The reception of the signal in the frame K' takes place at one point O' and the transition to the proper-time interval can be accomplished in accordance with Eq. (3.9): $T_0' = (1/\Gamma) T'$, so that the period T_0' will be equal for the observer at O' to

$$T_0' = \frac{1}{1 - B} T \sqrt{1 - B^2} = \sqrt{\frac{1 + B}{1 - B}}\, T. \qquad (3.11)$$

Passing over to frequencies ($\omega_0 = 2\pi/T$, $\omega' = 2\pi/T_0'$), we obtain

$$\omega' = \sqrt{\frac{1 - B}{1 + B}}\, \omega_0 \approx \omega_0 (1 - B). \qquad (3.12)$$

The last operation is performed in the easier way as follows: the numerator and denominator under the radical sign are multiplied by $(1 - \mathrm{B})$. The right-hand side of Eq. (3.12) is immediately obtained when the term B^2 is ignored in the denominator expression $(1 - \mathrm{B}^2)$. This equation is valid for an observer moving away from the source: the observed frequency is less than the proper one. When an observer approaches the source (in Fig. 3.5*b* the point O'' is to the left of O) and the signals are sent from the frame K'', the analogous reasoning (the source and the observer converge and the relative velocity is $c + V$) brings about the equations $T'' = \frac{1}{1+\mathrm{B}} T$, $T_0'' = \sqrt{\frac{1-\mathrm{B}}{1+\mathrm{B}}}\, T$ and finally

$$\omega'' = \sqrt{\frac{1+\mathrm{B}}{1-\mathrm{B}}}\, \omega_0 \approx (1 + \mathrm{B})\, \omega_0. \tag{3.13}$$

Eqs. (3.12) and (3.13) (without approximations) are the exact relativistic equations describing the radial Doppler effect. They will be obtained later (§ 7.2) on the basis of strict relativistic equations. But the derivation cited here is faultless in terms of physics. At the same time it clearly shows that the Doppler effect is formed of two independent parts: (1) it is connected with a continuously changing distance between the observer and the source; (2) it is also connected with the transformation of time intervals between events on transition from one reference frame to another. The first factor does not pertain to the theory of relativity in the slightest degree. The radial Doppler effect follows qualitatively from the classical theory, with the corresponding equation being obtained from Eq. (3.10). There is nothing to change in Eq. (3.10) in terms of the classical theory, since time intervals in all reference frames are the same. The difference between the classical equation and the relativistic one is essential only to the order of magnitude of B^2. The last approximate equation in (3.12) just gives the classical expression obtained from Eq. (3.10). Inasmuch as the ratio B is determined by the relative velocity of the source and the observer, it is very small at least for macroscopic sources, and the Doppler effect is defined primarily by a variation of the distance between a source and an observer. However, there is a case of a zero relative velocity of a source and an observer, although the frames, in which the source and the observer are at rest, move relative to each other. This happens when the moving source is observed at the moment when its velocity is perpendicular to the observation direction (the line of vision) (Fig. 3.6*a*). At the moment of observation illustrated in Fig. 3.6*a* the distance between the source and the observer does not vary. Consequently, no Doppler effect is possible from a clas-

sical viewpoint. But in terms of the relativistic theory the period T_0' between signals in the frame K' is the proper-time interval and $\omega_0 = 2\pi/T_0'$. Having converted T_0' into the observer's time according to Eq. (3.9), i.e. $T = T_0'\sqrt{1 - B^2}$, we obtain the equation of the transverse Doppler effect:

$$\omega' = \sqrt{1 - B^2}\,\omega_0. \tag{3.14}$$

This is the equation of the second order with respect to B. The transverse effect is more difficult to observe than the radial one,

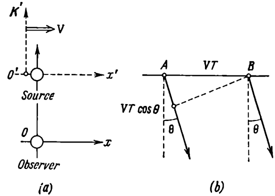


Fig. 3.6. The derivation of the Doppler effect equations: (*a*) the transverse effect; (*b*) the general case.

but still it was observed in 1938. Its discovery, as it is seen from the foregoing reasoning, is a direct evidence of the relativity of time intervals between events. It should be emphasized once more that the very existence of the transverse Doppler effect follows only from the STR. The equation of the Doppler effect can easily be derived if the radiation at the angle to the direction of the source motion is considered (Fig. 3.6*b*). If the first pulse is emitted at the point *A* and the second at the point *B*, the path difference of parallel rays travelling at the angle θ to the velocity direction is equal to $VT\cos\theta$. It is clear from this that $T' = T - \frac{VT\cos\theta}{c}$ in terms of the observer, and, consequently, $\omega' = \frac{\omega_0}{1 - \frac{V}{c}\cos\theta}$. Note that ω' and ω_0 in this equation are measured in the same reference frame. Thus we make sure once again that there is no transverse Doppler effect ($\theta = \pi/2$) in classical physics: $\omega' = \omega_0$.

We have made sure that when two events in a certain reference frame occur at the same point, i.e. are single-positioned, the time

interval between them is defined as the proper-time interval, i.e. is measured by one clock, whereas the time interval between the same events in any other frame can be calculated according to Eq. (3.9). The following question arises: is it always possible to convert a time interval found in an arbitrary reference frame into a proper-time interval? It turns out that this cannot be always done, and the stipulation, under which this becomes possible, will be found in § 3.4.

Now let us introduce the concept of the *object's proper time.* Let an object move uniformly and rectilinearly relative to the frame K. The frame K' can be fixed to the moving object. The object is at rest in this frame, so that events happening with this object or at it are registered by one clock. This clock counts the proper time at the point where the object is located; it can be said that this clock counts the object's proper time. Eq. (3.9) shows in this case that the interval between events that happened with the object or at it, is always less in terms of the object's proper time than the time interval between the same events registered by the clocks of any IFR relative to which this object moves. It should not be forgotten here that the proper-time interval is registered by one clock, while the time interval in the frame relative to which the object moves is registered by at least two clocks. This is very important because in interpreting Eq. (3.9) a moving clock is often said to have a slower rate than a stationary one. Such a mode of expression, however, may only confuse the situation. In fact, the clock rates are the same in all IFRs. What turns out to be different is the readings of time intervals between events. But it is only natural, because the clocks synchronized in one IFR are dis-synchronized in another.

The proper time can also be introduced for a particle moving with acceleration. To do this, let us consider the motion of a particle during an infinitesimal time interval. Let the particle velocity at a given moment be equal to V. Consider now the inertial frame of reference K^* moving at the velocity V. In this reference frame the equation $d\tau=\sqrt{1-\beta^2}dt$ is valid. This equation is also approximately valid for the instantaneously co-moving particle of the frame K'. The frame K^* differs from the instantaneously co-moving frame K' in that the latter, K', moves with acceleration, while the former, K^*, does not, although both of them move at the same velocity at a given moment. The less the time interval dt is, the more applicable becomes the equation $d\tau=\sqrt{1-\beta^2}\,dt$ in the frame K'. Having integrated this equation, we obtain the precise expression in terms of τ, which is in essence the overall "proper" time of any coordinate system K^*.

Proceeding from these considerations we shall suppose that if the interval between events that happened to the object turns out

to be equal to $d\tau$ when measured by the clock fixed to the reference frame co-moving with the object, the time interval between the same events dt, measured by the clock of another IFR relative to which the object moves, will be, according to Eq. (3.9),

$$dt = \gamma\, d\tau,$$

where

$$\gamma(t) = \frac{1}{\sqrt{1 - B^2}}, \quad \beta = \frac{v}{c}; \tag{3.15}$$

now $v = v(t)$ is the velocity of the object (and not of the reference frame). Owing to this circumstance the designations β and γ have been introduced. When the velocity of the object varies according to the equation $v = v(t)$ the relationship between the final time interval τ and the time interval registered by the clock of the frame relative to which the object moves is obtained by integration

$$\tau - \tau_0 = \int_{t_0}^{t} \sqrt{1 - \left(\frac{v(t)}{c}\right)^2}\, dt. \tag{3.16}$$

What is the quantity that is seen on the left-hand side of Eq. (3.16)? Of course, it can be called the object's proper time. But how to measure it? Strictly speaking, Eq. (3.9) is valid for the clocks in inertial frames of reference. But if the clock is fixed to an arbitrarily moving object, it will undergo an acceleration. No doubt, an acceleration affects the clock rate in a varying degree depending on the design of the clock. (If you do not believe in this, drop your clock on the floor.) Consequently, one can hardly speak of time readings made by means of such a clock. The reasonable interpretation of Eq. (3.16) lies in the fact that $\tau - \tau_0$ is the overall time measured in many inertial frames co-moving with the object or, which is the same, the time registered by the clock fixed rigidly to the object and *not affected at all by an acceleration of the object.*

It should be stressed that the difference in readings of clocks from different inertial frames of reference which we obtained has no relation whatever to any irregularity of the clock rate in one or another frame. As in the case of a measurement of a ruler's length, we deal here with different methods of time measurement. The rates of all clocks in all reference frames are absolutely the same. The measurements of time intervals between two events performed by the two sets of clocks from different reference frames, synchronized within their respective frames, lead to the result obtained: a proper-time interval between two events turns out to be always the least.

Let us consider an example, which is quite analogous to that discussed in connection with the variation of the scale's length, showing that deceleration of time is caused by different methods of its comparison. Let us take two quite identical clocks: A in the frame K and A' in the frame K'. These can be atoms of the same kind. Suppose we observe the clock A' of the frame K', i.e. compare the clock A with the set of clocks synchronized with the clock A'. Then observers from K' will discover that the clock A goes slower than the set of clocks from K'. On the contrary, observers from the frame K viewing one clock A' of the frame K' will discover that it goes slower than the set of clocks from K. Are these results contradictory? No. We clearly see that the methods of clock comparison in the first and the second cases are different. The clock which is compared with different clocks from another reference frame is always slower. This amazing situation proves to be inevitable. The equivalence of all IFRs underlies the theory of relativity, so when relative values emerge, they emerge in the same manner in all IFRs.

§ 3.4. **The classification of intervals and the principle of causality.** It is seen from Eqs. (3.1) and (3.2) that when two arbitrary events are considered, both the distance and the time interval between them prove to be relative values: $(\Delta x' \neq \Delta x,\ \Delta t' \neq \Delta t)$. Until now we dealt with the events of the special type: in length measurements the coordinates of a ruler's ends x_2 and x_1 were considered simultaneously $(t_1 = t_2)$; when a time interval was determined, the moments of time t_1' and t_2' were considered at the same point $x_2' = x_1'$. But even in those cases the spatial and temporal "distances" between events turned out to be relative. No wonder they do also in a general case. In addition to that, it follows directly from the Einstein postulates that the interval between events is, as we know, the invariant of the Lorentz transformation:

$$s_{12} = \sqrt{c^2 (t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2} = \\ = \sqrt{c^2 (\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2}. \quad (3.17)$$

The designations are the same as were used in the derivation of Eqs. (3.1) and (3.2). It is convenient to introduce also the special designations for the spatial and temporal distances between events:

$$l_{12}^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2, \quad \Delta t = t_{12}. \quad (3.18)$$

Having written out the squared interval between two events in the frame K as $s_{12}^2 = c_2 t_{12}^2 - l_{12}^2$ and in the frame K' as $s_{12}'^2 =$

$= c^2t'^2_{12} - l'^2_{12}$, we obtain the condition for the interval invariance $s'^2_{12} = s^2_{12}$ as

$$c^2t^2_{12} - l^2_{12} = c^2t'^2_{12} - l'^2_{12}. \tag{3.19}$$

Considering the events in an arbitrary reference frame K, we shall most likely discover that they happened at different points in space and at different moments of time.

Is it possible through the choice of the reference frame K' to ensure that (a) events I and II happen at the same point in space, i.e. be single-positioned; (b) events I and II happen at the same moment of time; and, finally, (c) events I and II happen at one point in space and at the same moment of time? Let us begin from the beginning.

(a) Is it possible to choose such a system K', in which these events will happen at the same point in space, i.e. will be single-positioned? This means that the following condition should be met: $l'_{12} = 0$. But then it follows from Eq. (3.19) that

$$s^2_{12} = c^2t^2_{12} - l^2_{12} = c^2t'^2_{12} \geqslant 0, \tag{3.20}$$

i. e. $s^2_{12} \geqslant 0$, and the interval s_{12} must be real. In the frame K' the events considered happen at one point in space and the time interval between them is equal (with an accuracy to within the factor c) to

$$t'_{12} = \frac{1}{c}\sqrt{c^2t^2_{12} - l^2_{12}} = \frac{s_{12}}{c}. \tag{3.21}$$

That is why the real intervals between events are referred to as *time-like intervals*. The condition for a time-like interval can also be written in the form $l_{12} < ct_{12}$.

Let us consider the motion of a particle possessing a rest mass. Conventional mechanics deals with objects of only this kind. Suppose for the sake of simplicity that this particle moves uniformly along the x axis, covering the distance Δx in the time Δt. In the frame K' this particle will travel the distance $\Delta x'$ in the time $\Delta t'$ which is determined in accordance with Eqs. (3.1) and (3.2). The ratio $\frac{\Delta x}{\Delta t} = v$ is the velocity of the particle in the frame K. Taking this into account, we can rewrite Eqs. (3.1) and (3.2):

$$\Delta x' = \Gamma(\Delta x - V\Delta t) = \Gamma\left(\frac{\Delta x}{\Delta t} - V\right)\Delta t = \Gamma(v - V)\Delta t \tag{3.22}$$

$$\Delta t' = \Gamma\left(\Delta t - \frac{V}{c^2}\Delta x\right) = \Gamma\left(1 - \frac{V}{c^2}\frac{\Delta x}{\Delta t}\right)\Delta t = \Gamma\left(1 - \frac{Vv}{c^2}\right)\Delta t. \tag{3.23}$$

Assuming $\Delta x' = 0$ in Eq. (3.22), one can easily find the velocity of the frame K' in which the two events in question are single-positioned. From the right-hand side of the equation we immediately obtain the evident answer $V = v$. So, this is just the frame co-moving with the particle. Another important consequence follows from Eq. (3.23). Let $\Delta t = t_2 - t_1 > 0$. This implies that event II happened after event I. Is there such a frame K' in which $\Delta t' < 0$, i.e. the time sequence of the events is inversed as compared to the frame K? Eq. (3.23) shows the sign of $\Delta t'$ to coincide with that of Δt when $\left(1 - \frac{Vv}{c^2}\right) > 0$. But this condition is always satisfied, since the velocity of an object v is always less than c. (The reference frame is also a material body.) The same condition is also satisfied for any pair of events coupled by a time-like interval. Indeed, the third link of Eq. (3.23) contains the expression $\left(1 - \frac{V}{c}\frac{\Delta x}{c\,\Delta t}\right)$. According to Eq. (3.18) $l_{12} \geqslant \Delta x$, and if $ct_{12} > l_{12}$, $c\,\Delta t > \Delta x$ *a fortiori*. This implies that the ratio $\Delta x/c\cdot\Delta t$ is less than unity; V/c is always less than unity, so that $\left(1 - \frac{V}{c}\frac{\Delta x}{c\,\Delta t}\right) > 0$ and, consequently, $\Delta t' > 0$.

Hence, for these two events, considered in terms of the frames K and K', the concepts "later" and "earlier" have an identical, that is absolute, character. In general, if the interval between events is time-like (recall that the interval is the invariant quantity), the time sequence of events remains the same throughout all IFRs. Later on we shall see that this is not the case for intervals which differ from time-like intervals by sign.

What is the significance of the invariance of time sequence in all inertial frames of reference? We have already ascertained that two events separated by a time-like interval will be single-positioned in some reference frame. If one of them happened "earlier" and the second "later", the first event may be the cause for the origination of the second, i.e. they may be connected by the cause-and-effect relationship. But in this case their time sequence cannot depend on the choice of a reference frame. It is from our results that the criterion for the possibility of the cause-and-effect relationship follows (the interval is time-like). As to the time sequence of events, it automatically remains the same in all reference frames.

The time-like interval between events indicates the possibility of a cause-and-effect relationship between events not only because it provides the identical time sequence in all IFRs. It indicates the physical opportunity of one event affecting another. It follows from the inequality $l_{12}^2 < c^2t_{12}^2$, determining a time-like interval, that during the time passing between the two events light can

a fortiori cover the distance from the point where event I occurred to the point where event II did, the product $c(t_2 - t_1)$ being the path travelled by light during the time $(t_2 - t_1)$. This means that basically a certain interaction (signal) could propagate from the point where event I occurred to the point where event II did during the time interval between the events. Without claiming the generality in formulating the problem we shall assume that one event can affect another only through a physical (force) interaction. Then, if event I happened, a "signal" about this fact can reach the point where event II will happen prior to the moment of occurrence of event II. This means that event I can be the cause of event II, and event II can be the effect of event I. In this case the events can have the cause-and-effect relationship. Thus, the events separated by a time-like interval can have the cause-and-effect relationship in terms of physics as well. It is understood that they may not be in such a relation. We only point to the theoretical possibility. What is essential, the time sequence cannot be upset in the case of such intervals: the consequence can never affect its cause.

(b) Now let us pass over to the consideration of intervals of the opposite sign. Let us examine again the condition of the interval invariance (Eq. (3.19)) and determine whether we can find such a coordinate system K' in which the two given events I and II happen simultaneously. This means that in this system $t'_{12} = 0$.

Hence $s'^2_{12} = -l'^2_{12} < 0$. The squared interval between the events must be negative, and the interval proves to be imaginary. In the frame K' the events in question happen at the same moment of time, and the interval between them is reduced (with an accuracy to within the number i) to the spatial interval $l'_{12} = is_{12}$. That is why imaginary intervals are referred to as *space-like intervals*. The condition for a space-like interval can also be written in the form $l_{12} > ct_{12}$.

Can one find the reference frame in which $\Delta t' = 0$ for two given events? Assuming $\Delta t' = 0$, we obtain from Eq. (3.2):

$$V = c\frac{c\,\Delta t}{\Delta x}. \tag{3.24}$$

Since one can always choose the events so that $l_{12} = \Delta x$, it follows from the condition for a space-like interval that in this case $\Delta x > c \cdot \Delta t$. Eq. (3.24) testifies that we can get $V < c$, that is, basically, such a frame can be chosen. The ratio $\Delta x / c \cdot \Delta t$ appears in the third link of Eq. (3.23); as we have mentioned, it can exceed unity. But this means that the factor $\left(1 - \frac{V}{c}\frac{\Delta x}{c\,\Delta t}\right)$ can be made negative by the appropriate choice of V.

It follows from here that the time sequence of two events related by a space-like interval can be reversed on transition from one IFR to another. This does not apply to the events which could have the cause-and-effect relationship. But in terms of physics they just cannot be so related. Indeed, the condition $l_{12}^2 > c^2t_{12}^2$ signifies that no "signal" can be transmitted from the point where event I happens to the point where event II does during the time interval between these events. Consequently, events separated by a space-like interval cannot be in a cause-and-effect relation.

Thus, the special theory of relativity makes it possible to indicate the conditions under which the cause-and-effect relationship becomes either possible or impossible. This is a very important criterion which cannot be obtained in the general form from other premises. It should be emphasized once more, of course, that all of our reasonings are based on the premise of the finite velocity of signal transmission.

(c) If we are interested in the reference frame in which the events would be both simultaneous and single-positioned, i.e. the two conditions $t'_{12} = 0$ and $l'_{12} = 0$ would be satisfied, the two inequalities $s_{12} \geqslant 0$ and $s_{12} \leqslant 0$ would have to be simultaneously complied with. This is possible only when $s_{12} = 0$ and $s'_{12} = 0$. If the events in question do not represent the sending and reception of light signals, the intervals can be equal to zero only in the case when the two events coincide in each of the frames K and K'. Of course, the coincidence of events does not depend on the choice of a reference frame.

The interval between two events that happened in a given frame at different points in space and at different moments of time and whose absolute value is equal to zero, is relevantly referred to as a *light-like interval*. A light-like interval links together events consisting in a light wave traversing consecutively various points in space. We made sure of this at the beginning of § 2.6.

§ 3.5. The transformation of velocity components of a particle on transition from one inertial frame of reference to another. From the Galilean transformation for coordinates and time (Eq. (1.2)) we obtained Eq. (1.4) showing how the particle velocity is transformed on transition from one IFR to another: $\boldsymbol{v}' = \boldsymbol{v} - \boldsymbol{V}$. This transformation rule does not satisfy the second postulate of Einstein, since the velocity of light *in vacuo* turns out to be different in different reference frames. The velocity transformation equations following from the Lorentz transformation satisfy the Einstein postulates. Now we proceed to the derivation of these equations.

Let us consider the particle motion in terms of the two IFRs: K and K'. The velocities are determined as usual:

In the frame K	In the frame K'
If	If
$x = x(t),\ y = y(t),\ z = z(t),$	$x' = x'(t'),\ y' = y'(t'),\ z' = z'(t'),$
then	then
$v_x = \frac{dx}{dt},\ v_y = \frac{dy}{dt},$ $v_z = \frac{dz}{dt}.$	$v'_x = \frac{dx'}{dt'},\ v'_y = \frac{dy'}{dt'},\ v'_z = \frac{dz'}{dt'}.$

The relationship between x, t and x', t' is meant to be established via the Lorentz transformation, so that t, for example, can be adopted as an independent variable. When t varies by dt, all variables get increments; differentiating the Lorentz transformation (see Eq. (2.16)), we obtain these increments in terms of differentials:

$$dx = \Gamma(dx' + V\,dt'),\quad dy = dy',\quad dz = dz',$$
$$dt = \Gamma\left(dt' + \frac{B}{c}\,dx'\right). \tag{3.25}$$

Having divided termwise the first three equations of Eq. (3.25) by the last one, we get

$$\frac{dx}{dt} = \frac{dx' + V\,dt'}{dt' + \frac{B}{c}dx'},\quad \frac{dy}{dt} = \frac{dy'}{\Gamma\left(dt' + \frac{B}{c}dx'\right)},$$
$$\frac{dz}{dt} = \frac{dz'}{\Gamma\left(dt' + \frac{B}{c}dx'\right)}.$$

Dividing the numerator and denominator of the right-hand sides of these equations by dt', we finally obtain

$$v_x = \frac{v'_x + V}{1 + \frac{V}{c^2}v'_x},\quad v_y = \frac{v'_y\sqrt{1 - \frac{V^2}{c^2}}}{1 + \frac{V}{c^2}v'_x},\quad v_z = \frac{v'_z\sqrt{1 - \frac{V^2}{c^2}}}{1 + \frac{V}{c^2}v'_x}. \tag{3.26}$$

These are *the formulae of the relativistic transformation of velocities*. Using them, the components v_x, v_y, v_z in the frame K can be found from the velocity components v'_x, v'_y, v'_z in the frame K' and the velocity of the frame K relative to K' which is equal to $-V$. In order to get the formulae for the inverse transformation, it is necessary to change the sign of the velocity V, to prime

the unprimed quantities and to withdraw primes from the primed ones:

$$v_x' = \frac{v_x - V}{1 - B\frac{v_x}{c}}, \quad v_y' = \frac{v_y\sqrt{1 - B^2}}{1 - B\frac{v_x}{c}}, v_z' = \frac{v_z\sqrt{1 - B^2}}{1 - B\frac{v_x}{c}}. \quad (3.27)$$

Of course, the same result will be obtained from Eq. (3.26) directly. It is seen from Eqs. (3.26) and (3.27) that a uniform motion in one IFR will be uniform in all other IFRs. Hence, the uniform and rectilinear motion is distinguished from all other kinds of motion. On the contrary, according to relativistic kinematics the uniformly accelerated motion in a certain IFR may not be that in other IFRs (see § 5.1).

In Eqs. (3.26) and (3.27) the x axis is distinguished from the y and z axes. This is only because the relative velocity of the reference frames K and K' is directed along the x axis. Passing to the limit $B \to 0$, i.e. assuming formally $c \to \infty$ in these equations, we get back to the Galilean transformation (Eq. (1.4)). This means that the Galilean transformation is sufficiently accurate if the relative velocity of the reference frames is slow compared to that of light. And hence we do not resort to relativistic concepts in our everyday life. Here we have learned again that the difference between relativistic and classical concepts is due to the finiteness of the velocity of light. Note that Eqs. (3.26) and (3.27) are casually derived using a four-dimensional approach to the theory of relativity.

Let us consider the motion along the x' axis. In this case the velocity components in the reference frame K' will be $v_x' = v'$, $v_y' = 0$, $v_z' = 0$. From Eq. (3.26) we see that in the frame K the components v_y and v_z are equal to zero. Consequently, the motion in the frame K also takes place along the x axis and $v_x = v$. That is why according to Eq. (3.26)

$$v = \frac{v' + V}{1 + \frac{V}{c^2}v'}. \quad (3.28)$$

Setting $v' = c$, we get $v = c$ from Eq. (3.28). This corresponds to the second postulate of Einstein: the velocity of light *in vacuo* is the same in all IFRs.

Note here that the substitution $v' = c$ in Eq. (3.28) is not quite consistent, since material particles representing a "signal" cannot move at the velocity c, the equation being derived for material particles $(m \neq 0)$. However, one may assume $v' = c$ in Eq. (3.28) considering light quanta (photons) as relativistic particles (§ 7.6). Besides, there are ultra-relativistic particles whose velocity is close to that of light. For example, the velocity of elec-

trons in the electron accelerator in Erevan differs from c in the eighth (!) decimal point.

We shall now give the explanation of the results of the Fizeau experiment as another example of utilizing Eq. (3.28) for the velocity summation. In this experiment the velocity of light propagating in water motionless relative to an observer (laboratory) was compared with that in water moving at the velocity V. The

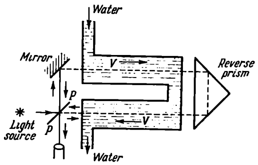


Fig. 3.7. The schematic drawing of the Fizeau experiment. A semi-transparent plate PP splits a light beam from a source into two light pencils, one going along and another against the water flow. Fresnel, who studied light propagation in moving media, anticipated the velocity of light $v = v' \pm kV$ for the observer relative to which water moves at the velocity V, with the velocity of light in motionless water being equal to v'; here the sign "+" corresponds to light propagation along the water flow, and the sign "—" to light propagation against the water flow; the coefficient k is called the drag coefficient. This coefficient was sought by Fizeau who conducted the described experiment. It followed from the Fizeau experiment that $k = 1 - 1/n^2$ The theory of relativity explains quite naturally Fizeau's result (see the text). The details of the experiment can be found in books [13] and [15].

velocity of light in motionless water is equal to c/n, where n is the refraction index of water. The light pencil travelled through the moving water (Fig. 3.7), and its velocity was determined in the laboratory frame of reference from an interference of two light pencils going one along and another against the water flow. It followed from the results obtained in the Fizeau experiment that the phase velocity of light in motionless water should be increased by the velocity of water V multiplied by $(1 - 1/n^2)$. Thus, if the phase velocity of light in motionless water is $v' = c/n$, the phase velocity, found in the laboratory frame, turns out to be

$$v = \frac{c}{n} + V\left(1 - \frac{1}{n^2}\right). \tag{3.29}$$

Using Eq. (3.28), we conclude that in the laboratory frame (due to the velocity summation law)

$$v = \frac{v' + V}{1 + \frac{V}{c^2} v'} = \frac{\frac{c}{n} + V}{1 + \frac{V}{cn}} = \frac{\left(\frac{c}{n} + V\right)\left(1 - \frac{V}{cn}\right)}{1 - \left(\frac{V}{cn}\right)^2}.$$

Neglecting the quantity $(V/cn)^2$ in the denominator whose smallness is due to the non-relativistic velocity of water ($(V/c)^2 \ll 1$), we obtain

$$v \approx \frac{c}{n} + V\left(1 - \frac{1}{n^2}\right) - \frac{V^2}{cn} = \frac{c}{n}\left[1 + \frac{V}{c}\left(1 - \frac{1}{n^2}\right) n - \frac{V^2}{c^2}\right]. \quad (3.30)$$

In Eq. (3.30) we again neglect the term V^2/c^2 and get the Fizeau result (Eq. (3.29)). Thus the Einstein equation for the velocity transformation provides a natural interpretation of the results of the Fizeau experiment (see § 1.7).

We have already pointed out that the most important assumption of contemporary physics is the statement concerning impossibility of transmitting signals (interactions) at the velocity exceeding that of light. No doubt, a moving object can be used to transmit a signal (energy, momentum), and, consequently, the velocity of an object cannot exceed c. Relativistic mechanics infers that the velocity of a material object, i.e. an object possessing a rest mass, is always less than c and never reaches this value. But this is valid in a definite IFR. Is it possible to choose such an IFR in which the velocity of an object will exceed c?

Had it followed from classical mechanics that in a given IFR the velocity of an object never exceeds c, one would have obtained the velocity of an object exceeding c via a choice of a suitable reference frame. Indeed, according to Eq. (1.4) $v = v' + V$, where v' is the velocity of the object relative to the frame K' and V is the relative velocity of the frames K and K'. If the velocities v' and V exceed 0.5 c, the velocity v of the object in the frame K will be greater than c.

But in the STR the velocity transformation is carried out differently. It is seen from Eqs. (3.26) and (3.27) that the velocities of a particle and of a frame do not add up as vectors do. Moreover, the velocity summation in the STR obeys the incredible rule $c + c = c$.

It follows from Eqs. (3.26) and (3.27) that if the velocity of a particle is less than c in the frame K ($v/c < 1$), and the velocity of the frame K' relative to K is also less than c ($V/c < 1$), the velocity of this particle determined in the frame K' is always less than c. The simplest demonstration of this statement can be carried out for the case of a unidimensional motion by means of Eq.

(3.28). Having composed the expression $(v/c - 1)$, we write out the following chain of equations:

$$\frac{v}{c} - 1 = \frac{\frac{v'}{c} + \frac{V}{c}}{1 + \frac{Vv'}{c^2}} - 1 = \frac{\frac{v'}{c} + \frac{V}{c} - 1 - \frac{Vv'}{c^2}}{1 + \frac{Vv'}{c^2}} =$$

$$= \frac{\frac{v'}{c}\left(1 - \frac{V}{c}\right) - \left(1 - \frac{V}{c}\right)}{1 + \frac{Vv'}{c^2}} = - \frac{\left(1 - \frac{V}{c}\right)\left(1 - \frac{v'}{c}\right)}{1 + \frac{Vv'}{c^2}} < 0. \quad (3.31)$$

Whence it is clear that $v < c$.

But is it possible to get the relative velocity of reference frames exceeding c by means of consecutive transitions from one frame to another? Strictly speaking, a reference frame is a system of material objects, so that in order to answer the question we can make use of the theorem just formulated. Certainly, in the STR one cannot obtain the relative velocity of frames exceeding c in any case. But now we shall derive this result once again by another method, which is instructive by itself.

Let us introduce, aside from the frame K, two more frames, K' and K''. What is the relative velocity of the frames K and K'' if, on the one hand, the relative velocity of the frames K' and K'', and, on the other hand, that of the frames K and K' are known? Let the relative velocity of K and K' be equal to V and that of K' and K'' be equal to W. Introducing the designations $\mathrm{B}_1 = V/c$ and $\mathrm{B}_2 = W/c$ and, correspondingly, $1/\Gamma_1 = \sqrt{1 - \mathrm{B}_1^2}$, $1/\Gamma_2 = \sqrt{1 - \mathrm{B}_2^2}$, we get

$$x = \Gamma_1 (x' + Vt'), \qquad t = \Gamma_1 \left(t' + \frac{\mathrm{B}_1}{c} x'\right), \quad (3.32)$$

$$x' = \Gamma_2 (x'' + Wt''), \quad t' = \Gamma_2 \left(t'' + \frac{\mathrm{B}_2}{c} x''\right). \quad (3.33)$$

Substituting Eq. (3.33) into Eq. (3.32), we find the explicit relationship between coordinates and time in the frames K and K''

$$x = \Gamma_1\Gamma_2 (x'' + Wt'' + Vt'' + \mathrm{B}_1\mathrm{B}_2 x'') =$$
$$= \Gamma_1\Gamma_2 \{(1 + \mathrm{B}_1\mathrm{B}_2) x'' + (V + W) t''\} =$$
$$= \Gamma_1\Gamma_2 (1 + \mathrm{B}_1\mathrm{B}_2) \left(x'' + \frac{V + W}{1 + \mathrm{B}_1\mathrm{B}_2} t''\right). \quad (3.34)$$

In much the same way one can obtain

$$t = \Gamma_1\Gamma_2 (1 + \mathrm{B}_1\mathrm{B}_2) \left(t'' + \frac{1}{c} \frac{V + W}{1 + \mathrm{B}_1\mathrm{B}_2} x''\right). \quad (3.35)$$

Designating

$$\frac{V+W}{1+B_1B_2}=\frac{V+W}{1+VW/c^2}=U, \tag{3.36}$$

we shall calculate the first multiplier in Eqs. (3.34) and (3.35):

$$\Gamma_1\Gamma_2(1+B_1B_2)=$$

$$=\frac{1}{\sqrt{\frac{(1-B_1^2)(1-B_2^2)}{(1+B_1B_2)^2}}}=\frac{1}{\sqrt{\frac{1+B_1^2B_2^2+2B_1B_2-(B_1^2+B_2^2+2B_1B_2)}{(1+B_1B_2)^2}}}=$$

$$=\frac{1}{\sqrt{1-\left(\frac{B_1+B_2}{1+B_1B_2}\right)^2}}=\frac{1}{\sqrt{1-\frac{U^2}{c^2}}}.$$

Then from Eqs. (3.34) and (3.35) one can get

$$x=\frac{x''+Ut''}{\sqrt{1-\frac{U^2}{c^2}}},\quad t=\frac{t''+\frac{U}{c}x''}{\sqrt{1-\frac{U^2}{c^2}}}.$$

Consequently, the two successive Lorentz transformations of the relative ~~velocities~~ V and W of the reference frames are equivalent to one transformation of the relative velocity U determined according to ~~Eq. (3.36).~~ In other words, the relative velocities of reference frames "add up" also according to Eq. (3.28). But we have already argued that such a summation will not produce a velocity greater than that of light.

Eq. (3.36) can readily be obtained by means of a complex rotation (see § 2.8). In geometrical terms the transition from K to K' and then from K' to K'' constitutes a consecutive rotation in the plane (x, τ) through the angles φ_1 and φ_2 with $\tan\varphi_1 = iB_1$ and $\tan\varphi_2 = iB_2$.

The tangent of the resulting angle can be found according to the conventional formula for the tangent of the sum of two angles ($\varphi = \varphi_1 + \varphi_2$):

$$\tan\varphi=\frac{\tan\varphi_1+\tan\varphi_2}{1-\tan\varphi_1\cdot\tan\varphi_2},$$

or

$$iB=\frac{iB_1+iB_2}{1+B_1B_2},$$

which is just Eq. (3.36) with B, B_1 and B_2 being replaced by their respective values. The two last expressions show that the set of the Lorentz transformations possesses the basic property of the group (from the standpoint of the mathematical group theory): two Lorentz transformations again produce a Lorentz transforma-

tion. It is essential here, however, that the relative velocity is always directed along the x axis.

Here is the useful interpretation of Eq. (3.28). In § 2.7 we introduced the parameter θ associated with the relative velocity of reference frames by the relationship $B = -\tanh\theta$. We can also introduce the velocity parameter for a particle $\beta = -\tanh\Theta$. Then Eq. (3.28) will take the following form:

$$\beta = \tanh\Theta = \frac{\beta' + B}{1 + \beta' B} = -\frac{\tanh\Theta' + \tanh\theta}{1 + \tanh\Theta' \cdot \tanh\theta} = -\tanh(\Theta' + \theta); \quad (3.37)$$

the last link of the equation is written in accordance with the equations of Appendix I, § 9. This is an interesting result. In the classical theory, it is velocities that add up (Eq. (2.4)), while in the relativistic theory velocity parameters do. The last contingency will be put to use in § 5.7.

§ 3.6. The transformation of an absolute value and the direction of the velocity of a particle. From Eq. (3.26) for the velocity component transformation one can obtain expressions determining the absolute value and direction of the velocity in the frame K, provided the velocity components in the frame K' are known. First of all, it is evident that if the velocity component $v'_z = 0$ in the frame K', the component $v_z = 0$ in the frame K too. This means that if the motion in the frame K' takes place in the plane (x', y'), the motion in the frame K will also take place in the plane (x, y). Let us choose the x' and y' axes so that the velocity of a particle lies in the plane (x', y') of the frame K'. Then it is clear that if θ' is the angle between the direction of the velocity v' and the x' axis then $v'_x = v'\cos\theta'$, $v'_y = v'\sin\theta'$. We shall denote the angle between the direction of the velocity v and the x axis in the frame K by θ. Consequently, $v_x = v\cos\theta$, $v_y = v\sin\theta$. Let us find the equations relating v and θ with v' and θ'. Having expressed the components v'_x and v'_y in terms of v' and θ', we can rewrite the first two formulae of Eq. (3.26) in the following form:

$$v\cos\theta = \frac{v'\cos\theta' + V}{1 + \frac{v'\cos\theta'}{c} B}, \quad v\sin\theta = \frac{v'\sin\theta'\sqrt{1 - B^2}}{1 + \frac{v'\cos\theta'}{c} B}. \quad (3.38)$$

Having divided the second formula by the first one, we obtain the expression for $\tan\theta$:

$$\tan\theta = \frac{v'\sqrt{1 - B^2}\sin\theta'}{v'\cos\theta' + V}. \quad (3.39)$$

In order to find the expression for the absolute value of the velocity, it is sufficient to square and sum termwise Eq. (3.38); we

shall obtain at once

$$v^2 = \frac{v'^2 + V^2 + 2v'V\cos\theta' - v'^2\mathrm{B}^2\sin^2\theta'}{\left(1+\frac{v'\cos\theta'}{c}\mathrm{B}\right)^2} = \frac{(v'+V)^2 - \frac{1}{c^2}[v'V]^2}{\left(1+\frac{v'V}{c^2}\right)^2}. \quad (3.40)$$

It follows from Eqs. (3.39) and (3.40) that the angle between the velocity direction and the corresponding x axis, as well as the absolute value of the velocity, change on transition from the frame K' to the frame K (Fig. 3.8a). (Recall that the geometric axes x and x' coincide.) Of course, the same occurs in classical mechanics as well, although it is described by other equations.

Now let us derive a useful formula resulting from Eq. (3.40). We shall need it when studying Chapter 7. Using the first equation of (3.40), we compose the following expression:

$$1 - \frac{v^2}{c^2} = \frac{c^2\left(1+\frac{v'V}{c^2}\cos\theta'\right)^2 - v'^2 - V^2 - 2v'V\cos\theta' + \frac{V^2v'^2}{c^2}\sin^2\theta'}{c^2\left(1+\frac{v'V}{c^2}\cos\theta'\right)^2} =$$

$$= \frac{c^2 - v'^2 + \frac{V^2}{c^2}v'^2 - V^2}{c^2\left(1+\frac{v'V}{c^2}\cos\theta'\right)^2} = \frac{\left(1-\frac{v'^2}{c^2}\right)\left(1-\frac{V^2}{c^2}\right)}{\left(1+\frac{v'V}{c^2}\cos\theta'\right)^2}.$$

Consequently,

$$\sqrt{1-\frac{v^2}{c^2}} = \frac{\sqrt{1-\frac{v'^2}{c^2}}\sqrt{1-\frac{V^2}{c^2}}}{1+\frac{v'V}{c^2}\cos\theta'}, \quad (3.41)$$

or

$$\sqrt{1-\beta^2} = \frac{\sqrt{1-\beta'^2}\sqrt{1-\mathrm{B}^2}}{1+\beta'\mathrm{B}\cos\theta'}. \quad (3.42)$$

From Eq. (3.41) it is also easy to obtain a convenient equation for the square of the absolute value of the velocity:

$$v^2 = c^2 - \frac{\left(1-\frac{v'^2}{c^2}\right)\left(1-\frac{V^2}{c^2}\right)}{\left(1+\frac{v_xV}{c^2}\right)^2}, \quad (3.43)$$

from which it immediately follows that if v'/c and V/c are less than unity, then $v < c$. For the special case, when the velocity was determined according to Eq. (3.28), the same theorem was demonstrated above. Eq. (3.43) can be obtained, of course, by squaring and summing up the left-hand and right-hand sides of Eq. (3.26).

From Eqs. (3.38) and (3.39) one can readily obtain the equations determining the variation of the direction of light rays on transition from the frame K' to the frame K. In this case, having assumed $v' = c$ in Eq. (3.43), we obtain, as it should be expected,

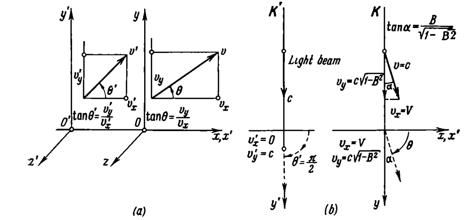


Fig. 3.8. (*a*) A particle moves in the plane (x', y') in the frame K'. The inclination angle of its velocity to the x' axis is equal to θ', with $\tan\theta' = v'_y/v'_x$. In the frame K the components v_y and v_x vary according to Eq (3 26), whence it is clear that the angle θ is not longer equal to θ' (see also Eq (3 39)). For the case shown in the diagram, $\theta' > \theta$. (*b*) Light propagates along the y' axis in the frame K', i. e. along the perpendicular to the frame motion direction. Obviously, $v'_x = 0$, $v'_y = c$, $\theta' = \pi/2$. In the frame K according to Eq (3 26), $v_x = V$, $v_y = c\sqrt{1 - B^2}$, whence $\tan\theta = \sqrt{1 - B^2}/B$. The aberration angle is formed by the visible direction of incoming light in the frame K and the direction of light in the frame K', i e. the y axis. The aberration angle $\alpha = \pi/2 - \theta$.

$v = v' = c$. Taking this into account, we obtain from Eqs. (3.39) and (3.38) respectively

$$\tan\theta = \frac{\sqrt{1 - B^2}}{B + \cos\theta'} \sin\theta', \tag{3.44}$$

$$\sin\theta = \frac{\sqrt{1 - B^2}}{1 + B\cos\theta'} \sin\theta', \quad \cos\theta = \frac{\cos\theta' + B}{1 + B\cos\theta'}. \tag{3.45}$$

Eqs. (3.45) describe the light aberration consisting in the wave front of a light wave changing its direction on transition from one IFR to another. Let light propagate in the frame K' along the perpendicular to the frame motion direction, for example, along the y' axis (Fig. 3.8*b*); this means that $\theta' = \pi/2$. Then, according to Eq. (3.44) $\tan\theta = \frac{\sqrt{1 - B^2}}{B}$.

The aberration angle is constituted by the visible directions of light in the two IFRs. In the frame K' light propagates along the

y' axis, while in the frame K at the angle $\alpha = \pi/2 - \theta$ to the y axis. Obviously, the angle α is the aberration angle, and

$$\tan\alpha = \tan\left(\frac{\pi}{2} - \theta\right) = \cot\theta = \frac{B}{\sqrt{1 - B^2}}.$$

The calculation of the aberration angle α according to the Galilean transformation (do the calculation yourself!) gives $\tan\alpha = B$. This means that the relativistic formula differs from the non-relativistic one by the term of the order of B^3.

One can also readily obtain the equations for the aberration angle in the case when light falls at an arbitrary angle to the motion direction. We shall assume that $B \ll 1$. Then, according to Eq. (3.45),

$$\sin\theta = \left(1 - \frac{1}{2}B^2 + \ldots\right)(1 - B\cos\theta' + \ldots)\sin\theta'.$$

Rejecting all the terms starting from B^2 and higher, we obtain

$$\sin\theta - \sin\theta' = -B\sin\theta'\cos\theta'.$$

The angle $\theta' - \theta = \Delta\theta$ is the aberration angle. Since the right part, proportional to B, is small, $\Delta\theta$ is small as well:

$$\sin\theta - \sin\theta' = 2\cos\frac{\theta + \theta'}{2}\sin\frac{\theta - \theta'}{2} \approx -\cos\theta'\Delta\theta.$$

Consequently,

$$\Delta\theta = B\sin\theta'. \tag{3.46}$$

This elementary equation describes the aberration of light falling at the angle θ' in the frame K'. Fig. 3.8 illustrates the change of the direction of particle velocity on transition from the frame K' to K, as well as the calculation of an aberration angle in the case of a perpendicular (relative to the motion) incidence of light. See Supplement II about the role that the aberration played in the development of the STR.

Finally, let us calculate the relative velocity of two particles. It is natural to define the relative velocity of two particles as the velocity of one of them in the frame K in which another particle is at rest. Let the velocities of the particles in the frame K' be $\boldsymbol{v}_1'$ and $\boldsymbol{v}_2'$. Choose the coordinate system K such that $\boldsymbol{V} = -\boldsymbol{v}_2'$. The particle velocities are immediately determined from Eq. (3.40). The absolute value of the velocity $\boldsymbol{v}_2$ is equal to zero, while that of the first particle is

$$\boldsymbol{v}_1^2 = \frac{(\boldsymbol{v}_1' - \boldsymbol{v}_2')^2 - \frac{1}{c^2}\left[\boldsymbol{v}_1'\boldsymbol{v}_2'\right]^2}{\left(1 + \frac{\boldsymbol{v}_1'\boldsymbol{v}_2'}{c^2}\right)^2}. \tag{3.47}$$

This expression defines the square of the relative velocity of the two particles. Eq. (3.47) is symmetric relative to v_1 and v_2.

It was shown (see Eq. (3.43)) that v^2 is always less than c^2 in Eq. (3.40). This is also the case for Eq. (3.47): the relative velocity of particles cannot exceed the velocity of light *in vacuo*.

§ 3.7. The K calculus (the radar method). We shall present below an elegant method of deriving the basic consequences of the Einstein postulates. This method could be briefly presented on the basis of the results obtained earlier by other methods. However, we intend to reiterate some conclusions in order to make this section more or less self-consistent. The method is remarkable because it dispenses with the coordinate routine, even in the derivation of the Lorentz transformation. Although graphical illustrations used below involve coordinates, they have an auxiliary character: these coordinates are not indispensable for the presentation of the method but are useful for those who are familiar with space-time diagrams.

Only one spatial coordinate is to be considered. Many characteristics of the STR are revealed even in this case permitting of descriptive illustrations. Thus, let all events occur on the x axis (and respectively on the coincident x' axis of the frame K', see Fig. 1.2). In the K calculus all conclusions are drawn from the imaginary experiments consisting primarily in the exchange of light signals *in vacuo*; their sending, reflection and reception are examined. In the final analysis, such a play with light spots makes it possible to obtain the basic consequences of the Einstein postulates.

The principal assumption which is made in the K calculus is based on the Doppler effect (see § 3.3); in a unidimensional case it is always the radial Doppler effect. Thus, if a stationary radar located in the frame K emits short pulses periodically with time intervals (periods) T, an observer in the frame K' moving away from this radar at constant velocity will discover that the interval between the incoming light pulses is different, despite the fact that the rates of the clock fixed to the radar and that of the observer from the frame K' are identical.

For the sake of simplicity we shall speak not of the radar and receiver, but of the two observers A and A' at rest in the reference frames K and K' respectively. Thus, if the observer A sends light signals separated by the time interval T according to his clock, the observer A' will receive these signals separated by a different interval as measured by his own clock. Let us designate this interval by KT. That is how the coefficient K appears, the key quantity of the considered method.

It should be pointed out that T and KT are the time intervals between the sending of the first and of the second signals by the

observer A and the reception of these signals by the observer A', measured in each case by the clocks at rest in the frames K and K' respectively.

Proceeding from the principal properties of space and time, their uniformity and isotropy, one can assume that the coefficient K depends neither on the positions of the receiver and the source, nor on the time of sending and receiving the signal, nor on the direction in which the signal is sent (in other words, the direction of the common x, x' axis may be chosen arbitrarily in space). Certainly, this coefficient does not depend on the time interval between the sendings of the signals. It may depend only on the relative velocity of the observers A and A'. Indeed, as the experience shows, the variation of the light frequency due to the Doppler effect depends only on the velocity of the relative motion.

The reason for the appearance of the coefficient K is evident. Let the observer A located at the origin of the reference frame K send light signals to the observer A' located at the origin of the frame K'. The frame K' moves away from the frame K to the right. Let the first signal be sent at the moment of time t. Then it is easy to determine the moment τ_1 by the clock of the observer A, when the observer A' receives this signal. Indeed, the signal propagating at the velocity c has to travel during the time τ_1 the distance Vt which separated the observers A and A' at the moment t and the distance $V\tau_1$ which will be covered by the observer A' during the time τ_1: $c\tau_1 = Vt + V\tau_1$, whence it follows that $\tau_1 = \frac{V}{c-V}t$. The second signal is emitted at the moment $t+T$ and it reaches A' in the time τ_2 determined from the equation $c\tau_2 = V(t+T) + V\tau_2$. Consequently, $\tau_2 = \frac{V(t+T)}{c-V}$. The difference $\tau_2 - \tau_1 = \frac{V}{c-V}T$ gives the time interval between the signals received by the observer A'. However, we have not yet found the expression for the coefficient K, although it may seem so. The coefficient K will be obtained as soon as we find the time interval between the incoming signals registered by the clock of the *observer A'*. But we have not determined the relationship between the readings of the clocks A and A' so far.

Until now we put to use only the uniformity and isotropy of time and space. Now we shall make use of the constancy of the velocity of light *in vacuo* in all IFRs. We shall have to use this property of light very often. This condition can be formulated as follows: "light cannot overtake light." Now we pass over to the problem which employs explicitly the equivalence of all inertial observers, i.e. the first postulate of Einstein.

We have agreed that the signals sent by the observer A at the intervals T will be received by the observer A' at the intervals KT

as measured by his clock. Due to the equivalence of the observers we have to suppose that the signals sent by the observer A' at the intervals T will be received by the observer A at the intervals KT as well. (The principle of relativity for two inertial observers A and A'.)

It is worth mentioning that this assumption is strongly based on the fact that vacuum contains no medium in which light propagates. Had such medium existed, the coefficient K would have depended on the velocities of the observers A and A' relative to this medium. It was just such a medium (ether) that agitated the minds of the 19th century physicists most of all. It caused a series of dramatic situations preceding the advent of the STR (see Supplement II). At present it is quite reasonable to adopt the contemporary point of view.

Now we shall find the explicit expression of the coefficient K in terms of a relative motion. In this procedure we shall need nothing except a few imaginary experiments pertaining to the sending, reflection and reception of light signals. The reflection can be treated, if necessary, as the sending of the signals by the "observer" in the reverse direction at the moment when he receives the incoming signal.

Let the first signal from the observer A to the observer A' be sent at the moment when the frames K and K' coincide. The observers A and A' located at the origins of their respective frames are positioned at this moment at the same point in space. Naturally, the transmission of this signal from A to A' and of the reverse signal from A' to A does not require any time. After the time interval T by his clock the observer A sends a light signal to the observer A' who will receive it in the time interval KT after the reception of the first signal. Let the observer A' send a signal back to A immediately on the reception of the second signal (the same as the mirror reflection). The two signals are separated by the time interval KT by the clock of the observer A'. Hence, the return signal will be sent from A' to A after this time interval. But the observer A will not receive it after the time interval KT. This time the interval will be increased K times again and will be equal to K^2T. Consequently, the return signal will be received at the moment K^2T by the clock of the observer A. Hence, in terms of the observer A the total travel of the second signal sent at the moment T to the observer A' and back takes the time $K^2T - T = (K^2 - 1)T$. Since the velocity of light is the same whether it propagates in the direct or the opposite direction, the propagation time from A to A' (or back) is equal to $^1/_2(K^2 - 1)T$. From this it follows that the determination of the distance between A and A' at the moment of reflection by means of a radar will give the value $^1/_2(K^2 - 1)\ Tc$.

Thus, we have found the distance between the observers A and A' at the moment when the signal is reflected. But at what moment by the clock of A did the reflection occur? Note that we speak of the clock located at A, while the event that we consider, i.e. the reflection of the signal at A', is removed from A. In this case we cannot measure the time of the event directly but have to *ascribe* a definite moment of time to it.

The second light signal was sent at the moment T and was received back at the moment K^2T. Hence, the moment of reflection is determined as $^1/_2(T + K^2T) = {}^1/_2(K^2 + 1)\,T$. Consequently, during the time interval $^1/_2(K^2 + 1)\,T$ the observer A' moves away from the observer A by the distance $^1/_2(K^2 - 1)\,Tc$. So, the relative velocity of the observer A' is

$$V = \frac{\frac{1}{2}(K^2 - 1)\,Tc}{\frac{1}{2}(K^2 + 1)\,T}, \quad \text{or} \quad \frac{V}{c} = \frac{K^2 - 1}{K^2 + 1}. \tag{3.48}$$

It therefore follows * that

$$K = \sqrt{\frac{1 + B}{1 - B}}. \tag{3.49}$$

Here we shall write the two equations which we shall need later:

$$\frac{K^2 + 1}{2K} = \frac{1}{\sqrt{1 - B^2}} \equiv \Gamma, \quad \frac{K^2 - 1}{2K} = \Gamma B. \tag{3.50}$$

It is very convenient to make use of a graphical diagram in order to present descriptively the results obtained. Let us introduce the Cartesian coordinate system on the plane with coordinate axes x and ct. Later we shall see that the choice of the spatial and time coordinates of the same dimension is downright inevitable, but for the present we shall be marking τ along the axis of ordinates which is proportional to time: $\tau = ct$. The x and τ axes are drawn in Fig. 3.9. Every point of the plane represents the event defined by the coordinates (x, τ). The motion of a body is a sequence of events consisting in the arrival of this body at a given point at a given moment of time; it is depicted as a curve in the plane (x, τ).

The uniform motion of an object is depicted in this plane by a straight line. The propagation of a light beam at the velocity c is depicted by a bisecting line (the equation $x = \tau$) running through quadrants I and III when light propagates in the positive direction of the x axis and through quadrants II and IV when light

* Later we shall see that this equation determines the change of a light frequency on reflection from a moving mirror (see § 7.5).

travels in the opposite direction. Since the velocity of an object is always less than that of light, the uniform motion of any object is depicted by a straight line forming an angle less than $\pi/4$ with the τ axis.

It is easy to find the points on the plane which depict the motion of the observers A and A'. In the frame K shown in Fig. 3.9 the observer A is at rest; we shall suppose that he is located at the point $x = 0$. Then his "world line", i.e. the succession of points in the plane (x, τ) corresponding to the events consisting in his being at a given point at a given moment of time, will be represented by the axis of ordinates. Hence, the axis of ordinates is the world line of the observer A. The world line of the observer A' in the frame K is represented by the straight line inclined to the τ axis at the angle α whose tangent is determined by the ratio $\tan\alpha = x/\tau = x/ct = v/c$. If at the moment $t = 0$ the observers A and A' were located at one point, the world line of the observer A' passed through the origin O. The sending and reception of light signals by the observer A is depicted in the plot (x, τ) as follows. The first "exchange" of signals takes place at the point O. Then, after the time interval T (at the world point A_1), the observer A sends a light signal. Its propagation is described by the straight line A_1A_1' parallel to the bisecting line. The observer A' will receive the light signal at the world point A_1'. The propagation of the light signal sent by the observer A' in the reverse direction is depicted by the straight line $A_1'A_2$ parallel to the bisecting line of quadrants II and IV not shown in Fig. 3.9. The observer A will receive the return signal at the world point A_2. According to the condition $OA_1 = T$ * and, from the definition of the coefficient K, $OA_2 = K^2T$. In the frame K the point A_1' is associated with the moment of time (by the clock located at the point $x = 0$, i.e. at the observer A) A_3. Obviously, $OA_3 = {}^1/_2(OA_1 + OA_2) =$

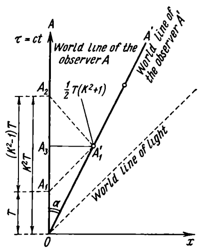


Fig. 3.9. The graphic illustration of the determination of the relative velocity of two observers.

* In the coordinates τ one should write $OA_1 = cT$, but for the sake of simplicity we shall not do this.

$= {}^1/_2(K^2+1)\ T$. The propagation time of the second signal from A to A' is, naturally, ${}^1/_2(OA_2 - OA_1) = {}^1/_2(K^2-1)\ T$.

Next, we shall note a useful theorem of the K calculus. It is seen from Eq. (3.49) that the change of the sign of the relative velocity, i.e. of the quantity B, transforms the quantity K into $1/K$. This means that the receding and the approaching at the same absolute value of the velocity correspond to reciprocal values of the coefficient K.

Let us now consider the case when there are three reference frames K, K' and K'' and three observers located at the corresponding origins O, O' and O''. Let the coefficient K be equal to $K(A, A')$ for the observers A and A'; it depends only on the relative velocity of the frames K and K' which we shall designate by V as before. If the relative velocity of the observers A' and A'' is equal to W, the coefficient K for these observers, $K(A', A'')$, depends only on W. Is it possible to find $K(A, A'')$ when $K(A, A')$ and $K(A', A'')$ are known? Let us derive the requisite equation.

Let the observer A send two light signals separated by the time interval T registered by his clock. The observer A' receiving these signals will find that they come in separated by the time interval $K(A, A')\ T$ as it follows from the definition of the coefficient K. But this time is registered by the clock of the observer A'. The observer A'' is located further from A than the observer A', so that the signals passing the observer A' go onward to A''. At the moment when A' receives the first signal from A, he sends a light signal himself without delay to A'' (do not worry: it is an imaginary experiment!). Now two signals propagate toward the observer A'': one travelling from A and another sent by the observer A'. Since both of them are light signals, they propagate at the same velocity, having left A' at the same moment. In fact, they propagate as one signal.

The same procedure is repeated by the observer A' at the moment when the second signal from A comes in. And again one signal propagates from A' to A'', consisting of two light pulses sent from A and from A'.

The observer A'' will receive the two signals. On the one hand, according to the definition he will register by his clock that the time interval between the signals is equal to $K(A, A'')\ T$. On the other hand, these signals were sent by the observer A' with the time interval between them $K(A, A')\ T$. According to the definition the observer A'' will find that the time interval between these signals is equal to $K(A', A'') \cdot K(A, A')\ T$. But the signals from A, and A' arrive at A'' simultaneously, so that

$$K(A, A'') = K(A, A') \cdot K(A', A''). \tag{3.51}$$

The result is remarkably simple. Knowing the coefficients K for two pairs of reference frames in which one common frame is contained, one can obtain the unknown coefficient K pertaining to the last pair of frames by multiplication of the known coefficients K.

Graphically this result is readily obtained from Fig. 3.10. Here the world lines of the three observers A, A' and A'' are depicted. All the observers were at the same point O at the moment $t = 0$. The time interval T later, the observer A sends from the world point A_1 a light signal whose world line is depicted by a dotted straight line $A_1A_1'A_1''$. According to the condition, $OA_1 = T$ and by the definition $OA_1' = K(A, A')T$, $OA_1'' = K(A, A'')T$. On the other hand, evidently $OA_1'' = K(A', A'') \cdot K(A, A_1')T$. Note that the proposed diagram is suitable only for a graphical depiction of "imaginary experiments" but cannot be used for a geometric determination of various quantities. The plane (x, τ) is not just a conventional Euclidean plane (see Chapter 4). However, combining a graphical geometric description with algebraic determinations, we shall not make a mistake.

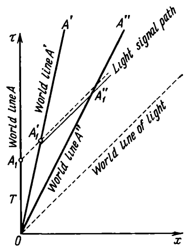


Fig. 3.10. The derivation of Eq. (3.51).

It is easy to find the equation for the transformation of velocities of coordinate systems. Suppose we want to find the relative velocity U of the frames K and K'' if the relative velocity of the frames K and K', designated by V, and that of the frames K' and K'', designated by W, are known.

Introducing the familiar designations $V/c = B_1$ and $W/c = B_2$, we obtain from Eqs. (3.48), (3.49) and (3.51)

$$\frac{U}{c} = \frac{K^2 - 1}{K^2 + 1} = \frac{K^2(A, A') \cdot K^2(A', A'') - 1}{K^2(A, A') \cdot K^2(A', A'') + 1} = \frac{B_1 + B_2}{1 + B_1B_2}.$$

Going back to conventional designations, we obtain the equation for the velocity transformation (Eq. (3.28)):

$$U = \frac{V + W}{1 + VW/c^2}.$$

It is seen from the reasoning quoted that the moments of occurrence of events and the time intervals between them prove

to be different for observers from different IFRs. To detect this, let us return to the experiment analysed earlier and involving an exchange of light signals between the observers A and A'. We shall recall that the first "exchange" is performed at the moment when the observers are located at one point. At that very moment the clocks of the observers A and A' are set to the zero reading. Then after the time interval T by his clock the observer A sends a signal directed to A'; according to the definition, the time interval separating the reception of the first and the second signals by the observer A' is equal to KT by his clock. However, the observer A will *ascribe* the moment of time $^1/_2(K^2+1)\,T$ to the reception of the signal at A' and will assume that the signals sent by him at the intervals T will reach A' with the intervals $^1/_2(K^2+1)\,T$. As it was mentioned, the same interval in terms of the clock A' is equal to KT. Hence, the time interval between the two identical events, the arrival of the first and the second signal at A', proves to be different: in terms of A' it is equal to KT and in terms of A it is equal to $^1/_2(K^2+1)\,T$. Thus, we discovered that the time of the event, i.e. the arrival of the second signal, is *relative*: it is equal to KT in terms of A' and $^1/_2(K^2+1)\,T$ in terms of A. The time interval between the two events proved to be different for A and A' too. All this indicates that the time of an event as well as the time interval between events are relative values.

Under what conditions will these values coincide? It happens when $KT \approx {}^1/_2(K^2+1)\,T$. It can readily be inferred that it is possible when $K \approx 1$ or, as it is seen from Eq. (3.48), when $V/c \to 0$. Thus, the difference in time readings and the relativity of time intervals between events can be neglected in those IFRs whose relative velocities are small compared to that of light.

The proper time. The K calculus makes it possible to determine readily a relationship of a time interval between two events that occur in a certain IFR at one point in space and are, consequently, registered by one clock (the proper-time interval), and a time interval between the same events registered by two clocks of another IFR in which the considered events occur at different points.

Now let us go back to the exchange of light spots. If A sends signals at the interval T by his clock, A' receives them at the interval KT by his clock. However, as we saw before (p. 108), this interval is equal to $^1/_2(K^2+1)\,T$ in terms of A. It is the ratio of these quantities that gives the relationship between the proper-time interval $\Delta\tau = KT$ and the time interval Δt registered by two clocks of another IFR. This ratio is equal to

$$\frac{\Delta\tau}{\Delta t} = \frac{KT}{1/2\,(K^2+1)\,T} = \frac{2K}{K^2+1} = \sqrt{1-\frac{V^2}{c^2}}\,,$$

where in the last link Eq. (3.48) is used. This result is of course familiar to us.

Relativity of rulers' lengths (distances). Suppose we have two motionless points in the reference frame where the observer A' is at rest. One may presume, although it is far from being obligatory, that these points are a ruler's ends. Let the ruler move from the observer A and the observer A' be located at the end of the ruler which is nearer to A. (Do not forget that the ruler is oriented along the direction of the relative velocity.)

To determine the length of the ruler, the observer A sends a signal at the moment t_1, registered by his clock, and waits for it to return after reflection from the far end of the ruler. Let the moment of the signal return be t_4 by the clock of A. Obviously, the moment of the signal reflection is equal to $^1/_2(t_1 + t_4)$. Exactly in the same manner, a signal can be sent to the near end of the ruler (say, at the moment t_2) and the moment of its return determined (for example, t_3). The moment of the signal reflection from the near end is equal to $^1/_2(t_2 + t_3)$. Both signals are reflected simultaneously (by the clock of A) from both ends of the ruler, provided the following condition is met:

$$\frac{1}{2}(t_1 + t_4) = \frac{1}{2}(t_2 + t_3). \tag{3.52}$$

In an imaginary experiment this condition can be satisfied by choosing the times of sending of the first and the second signals.

The first signal from A, however, will be received by the observer A', located at the near end of the ruler, at the moment Kt_1 (recall that the initial readings of the clocks of A and A' coincided when the observers were located at one point). The signal reflected from the far end of the ruler and returning to A at the moment t_4 will pass A' at the moment t_4/K. Indeed, the signal received by A' at the moment t_4/K will get to the observer A at the moment $(t_4/K)\cdot K = t_4$. From the viewpoint of the observer A' the doubled length of the ruler l_0 is determined as the time interval, taken by light to reach the far end of the ruler and get back, multiplied by the velocity of light, i.e.

$$\frac{1}{2}\left(\frac{t_4}{K} - Kt_1\right)c = l_0 \tag{3.53}$$

As to the relationship between t_2 and t_3, it follows directly from the definition of the coefficient K:

$$t_3 = K^2 t_2. \tag{3.54}$$

The imaginary experiments performed to measure length are illustrated in Fig. 3.11, which does not require any special explanations after the diagrams of Figs. 3.9 and 3.10 have been analysed.

The ruler's length determined by the observer A is equal to the difference of the distances from him to the far and near ends of the ruler under the necessary condition that these distances are determined simultaneously. This condition is satisfied owing to the validity of Eq. (3.52). The distance from A to the far end is equal to $^1/_2(t_4 - t_1)\,c$ and to the near end to $^1/_2(t_3 - t_2)\,c$. Consequently, A has to assume the ruler's length l equal to

$$l = \frac{1}{2}\left[(t_4 - t_1) - (t_3 - t_2)\right] c. \tag{3.55}$$

Eqs. (3.53)-(3.55) make it possible to find the relationship between l and l_0. It follows from Eq. (3.52) that $t_4 = t_2 + t_3 - t_1$. Substituting the expression obtained for t_4 into the left-hand side of Eq. (3.53) and resorting to Eq. (3.54), we get

$$l_0 = \frac{c}{2}\left(\frac{t_2 + t_3 - t_1}{K} - Kt_1\right) = \frac{c}{2}\left[\frac{t_2(K^2+1) - t_1(K^2+1)}{K}\right] = \frac{c}{2}\,\frac{K^2+1}{K}(t_2 - t_1). \tag{3.56}$$

Since according to Eq. (3.52) $t_2 - t_1 = t_4 - t_3$, it follows from Eq. (3.55) that

$$t_2 - t_1 = \frac{(t_2 - t_1) + (t_4 - t_3)}{2} = \frac{(t_4 - t_1) - (t_3 - t_2)}{2} = \frac{l}{c}.$$

Now Eq. (3.56) takes the form

$$l_0 = l\,\frac{K^2+1}{2K} = \frac{l}{\sqrt{1 - B^2}},$$

where in the last equation the formula (3.50) is taken into account. This is exactly what we obtained earlier as Eq. (3.5).

This derivation shows quite distinctly how essential it is to find the ruler's ends simultaneously when its length is determined. Incidentally, note that the derivation of Eq. (2.4) involves, in essence, a radar approach as well.

The Lorentz transformation. We have made sure that the K calculus can be employed to derive all basic principles of the STR, the Einstein postulates. The advantage of this derivation lies in the fact that there is no need for an explicit introduction of a coordinate system.

But, of course, the application of STR methods in physics requires an explicit introduction of a reference frame. If so, the introduction of the Lorentz transformation is outright inevitable. The Lorentz transformation can be derived by means of the K calculus.

Consider the two reference frames K and K' with the respective observers A and A' registering the same event. In both frames the

initial time reading is chosen so that $t = t' = 0$ when both origins coincide. Then at the moment t_1 the observer A sends a light signal to A' which is received by him at the moment t_1' by his clock; the signal sent by A proceeds further accompanied by the signal sent by A' at the moment when he receives the signal from A. In fact, one signal consisting of two propagates along the x axis. Let the event P represent the arrival of that signal at some point (or the arrival of the signal coincides with the moment of occurrence of a certain event). At that

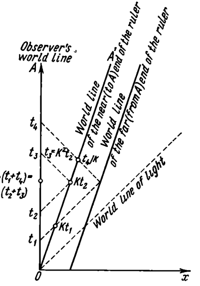


Fig. 3.11. The determination of the length of a moving ruler.

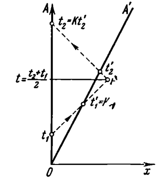


Fig. 3.12. The derivation of the Lorentz transformation.

point the signal is reflected (or, otherwise, the return signal is sent immediately on the arrival of the direct one). First, it gets to the observer A' at the moment t_2'; at the same moment A' sends his signal in the direction of A. Now the single signal consisting, in fact, of two signals propagates from A' to A. It is received by the observer A at the moment t_2 (Fig. 3.12).

The observer A will ascribe the coordinates to the event P as follows. The time t of the event is just the half-sum of the times of sending and reception of the signal since the velocity of light on the way "there" and "back" is equal:

$$t = \frac{1}{2}(t_1 + t_2). \tag{3.57}$$

The distance to the point where the event occurred can be found if the propagation velocity of the signal c is multiplied by the time which the signal takes to travel "there"; this time is equal to

half the total time spent by the signal. Since the signal travelled a closed path during the time $t_2 - t_1$, the event coordinate x will be determined by the observer A as

$$x = \frac{1}{2}(t_2 - t_1)c. \tag{3.58}$$

From Eqs. (3.57) and (3.58) we obtain

$$t_1 = t - \frac{x}{c}, \quad t_2 = t + \frac{x}{c}. \tag{3.59}$$

But the observer A' will find in exactly the same manner that

$$t'_1 = t' - \frac{x'}{c}, \quad t'_2 = t' + \frac{x'}{c}. \tag{3.60}$$

According to the definition of the coefficient K, and comparing the intervals between the exchanges of signals, we get

$$t'_1 - 0 = K(t_1 - 0), \quad t_2 - 0 = K(t'_2 - 0). \tag{3.61}$$

According to Eqs. (3.59) and (3.60) we obtain

$$t' - \frac{x'}{c} = K\left(t - \frac{x}{c}\right), \tag{3.62}$$

$$t + \frac{x}{c} = K\left(t' + \frac{x'}{c}\right). \tag{3.63}$$

Multiplying crosswise Eqs. (3.62) and (3.63), we immediately obtain that the quantity

$$t'^2 - \frac{x'^2}{c^2} = t^2 - \frac{x^2}{c^2}, \tag{3.64}$$

retains its value in all IFRs, i.e. is the invariant. Having written Eqs. (3.62) and (3.63) in the form more convenient for solution

$$t' - \frac{x'}{c} = K\left(t - \frac{x}{c}\right), \tag{3.65}$$

$$t' + \frac{x'}{c} = \frac{1}{K}\left(t + \frac{x}{c}\right), \tag{3.66}$$

we readily find that

$$t' = \frac{K^2+1}{2K}t - \frac{K^2-1}{2Kc}x, \quad x' = \frac{K^2+1}{2K}x - \frac{K^2-1}{2K}ct.$$

Taking into account Eq. (3.50), we discover that this is just the Lorentz transformation:

$$x' = \Gamma(x - Vt), \quad t' = \Gamma\left(t - \frac{V}{c^2}x\right).$$

CHAPTER 4

THE FOUR-DIMENSIONAL SPACE-TIME

§ 4.1. Three-dimensional and four-dimensional Euclidean spaces. When we introduce a coordinate system, the position of every point is specified by three numbers which are referred to as the coordinates of a point. A manifold of three dimensions is understood as a set of all points. If we want to pass over from a manifold to space possessing definite geometrical properties, we have to define the expression for a distance between two infinite close points of the manifold. Having assigned the square of the distance between such points, one can define basic geometric quantities, such as a vector's length, an angle between vectors, areas of two-dimensional figures formed by vectors. Geometry, whose principal laws were formulated by Euclid, is valid to a high degree of accuracy in the world that we live in. In accordance with Euclidean geometry the square of the distance between two infinitely close points can be put down in the Cartesian coordinates in the following form:

$$ds^2 = dx^2 + dy^2 + dz^2. \tag{4.1}$$

This equation represents nothing other than the Pythagorean theorem written out for the diagonal of a rectangular three-dimensional parallelepiped with the sides dx, dy, dz.

A coordinate system can be selected at will (the Cartesian coordinate system is distinguished only for its simplicity), and a distance between points, owing to its geometric meaning, should not depend on the choice of a system. This means that Eq. (4.1) has to be an invariant of any transformation of coordinates. A distance between any two points has also to be an invariant of the transformation of coordinates. Thus, in Euclidean geometry the invariant is the distance between two points:

$$r_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}, \tag{4.2}$$

where (x_1, y_1, z_1) and (x_2, y_2, z_2) are the coordinates of two points in space. Eqs. (4.1) and (4.2) relate the coordinates of two points of space. In particular, the transformation equations for transitions from one Cartesian system to another are given in Ap-

pendix 1, § 2. Such a transition represents a rotation, provided we ignore the system's translation which is of little interest to us. It is seen from this equation that in a new coordinate system any new coordinate is expressed through all old ones.

In a three-dimensional Euclidean space one can introduce vectors specified by a triad of numbers, i.e. vector components. The coordinates of a point comprise the components of a radius vector. Consequently, components of any vector are transformed according to the coordinate transformation rule. Norms of vectors, their dot products and an angle between them are found via vector components according to the known rules.

What would the appearance of one more dimension in a Euclidean space imply? Certainly, it is difficult to visualize a four-dimensional space with one's own eyes. But there is no such need. Having available the principal relationships for a three-dimensional space, we just carry them over to a four-dimensional space. Let the coordinates of a point in the four-dimensional space be x, y, z, w. For a four-dimensional Euclidean space the square of the distance between two infinitely close points will be written in the following form (the symmetrical designations are also given):

$$ds^2 = dx^2 + dy^2 + dz^2 + dw^2 = dx^{0^2} + dx^{1^2} + dx^{2^2} + dx^{3^2}, \quad (4.3)$$

and the distance between points

$$r_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 + (w_2 - w_1)^2}. \quad (4.4)$$

Eqs. (4.3) and (4.4) will be the invariants of the coordinate transformation, and basic geometrical relationships will be found in much the same way as they are found in three-dimensional space.

§ **4.2. The 4-space-time, or the four-dimensional pseudo-Euclidean space.** Let us consider a four-dimensional manifold made up of "points" whose coordinates are constituted by four numbers x, y, z, $\tau = ct$ defining a four-dimensional point. One or another event representing an instantaneous physical process can occur at any point of this manifold. The four-dimensional space-time is a purely geometric notion. Sometimes, following Minkowski, this space is called the "world". Any event occurs at some point of the Minkowski world.

Geometric properties of the Minkowski world can be established after some invariant relationship between coordinates of points is found, which can be interpreted as the distance between two points of a manifold. When the distances between points are defined, we may pass from manifold to space. But how can the necessary invariant relation be found? It should not be forgotten that the coordinates of the "world" points are defined with physi-

cally different quantities, so that it is impossible to presume in advance that the "distance" in this world can be defined by the expression of the (4.3) type. But the theory of relativity answers this question unambiguously. Considering only inertial frames of reference, the interval between events (Eq. (3.19)) remains the invariant for any pair of events, or, in terms of geometry, for any pair of points in the Minkowski world. The transition from one IFR to another is described by the Lorentz transformation, and no other transformation is needed in the framework of the STR. Consequently, from physical considerations we can take the expression for the square of the interval between events

$$ds^2 = d\tau^2 - dx^2 - dy^2 - dz^2 = dx^{0^2} - dx^{1^2} - dx^{2^2} - dx^{3^2} \quad (4.5)$$

as the basic invariant quadratic form defining the "distance" in the Minkowski world. Here $\tau = ct$. It is Eq. (4.5) that defines the square of the distance between two infinitely close points in the Minkowski world. Thus, the Einstein postulates, from which the invariance of the interval between events follows, signify that geometry of the four-dimensional space-time, i.e. the Minkowski space, is determined by the basic fundamental form of the (4.5) type. It is seen from the appearance of this form that coordinates and time are not equivalent.

It will be shown in Supplement V that the transition from inertial frames of reference to non-inertial ones alters the appearance of the interval between events. Although this expression always remains invariant, its form becomes different, so that the square of the interval takes the following form:

$$ds^2 = g_{ik}\, dx^i\, dx^k, \quad (4.6)$$

where the indices i and k denote summation from 1 to 4 and the coefficients g_{ik}, referred to as *metric coefficients*, may depend on coordinates and time. We need this general equation now only in order to write out g_{ik} for Eqs. (4.3) and (4.5). Using the symmetric designations of Eqs. (4.3) and (4.5), we obtain respectively

$$g_{00} = g_{11} = g_{22} = g_{33} = 1, \quad (4.3')$$

$$g_{00} = 1, \quad g_{11} = g_{22} = g_{33} = -1. \quad (4.5')$$

It is evident that Eqs. (4.3) and (4.5) differ by signs of metric coefficients. A totality of these signs is called a *signature* of corresponding quadratic forms. The signature of Eq. (4.3) has the form $(++++)$, while the signature of Eq. (4.5) $(+---)$. If a four-dimensional space were formed by a simple increase of the number of dimensions of our conventional space, the signature would be $(++++)$. Such a space would not differ from our space in anything except the number of dimensions and it would

be referred to as the Euclidean (four-dimensional) space. The signature of space considered in the special theory of relativity corresponds to that of Eq. (4.5), i.e. $(+---)$.

A change of a signature implies a variation of the "distance" between points in space, the variation of properties of this space as compared to those of the customary Euclidean space. This four-dimensional space, possessing unusual geometric properties, is extremely important for the STR. It is in this space that all physical phenomena take place.

The geometry of the Minkowski world differs from Euclidean geometry, but not too much, since the coefficients in Eq. (4.5) as well as those in Eq. (4.3) are constant. Accordingly, the geometry defined by the quadratic form of Eq. (4.5) is customarily called *pseudo-Euclidean* and the corresponding space a *pseudo-Euclidean space.*

Thus, the space of the four variables x, y, z, ct of the special theory of relativity is the four-dimensional pseudo-Euclidean space. It is not originated just by adding the fourth (time) coordinate ct to the three spatial ones x, y, z, but through the peculiar definition (Eq. (4.5)) of the invariant distance between the points of this space.

A physical motive for the consideration of the pseudo-Euclidean space lies in the fact that the spatial and time readings pertaining to an event are not equivalent in the STR in spite of their close connection.

§ 4.3. 4-vectors and 4-tensors. Exactly as in a three-dimensional space, coordinates of a point in a four-dimensional space can be treated as components of a four-dimensional radius vector drawn from the origin of a coordinate system to a given point. All four-dimensional vectors will be designated by an arrow over a letter; in particular, a four-dimensional radius vector will be designated by $\vec{R}$. For the convenience of our readers we shall be presenting basic relationships both in complex notation and via the real variables. Complex notation simplifies the presentation of electrodynamics, while the usage of real variables leads us to the formalism of the general theory of relativity, where the introduction of a complex coordinate is of no use. Most of the equations will be written in a symmetric notation, and the presentation of coordinates of a four-dimensional vector in a two-line form will make it possible to recall the meaning of the introduced designations. Thus, we introduce a four-dimensional radius vector in one of the following ways:

$$\vec{R}\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x & y & z & i\tau = ict \end{pmatrix}; \quad \text{(a)} \quad \Big| \quad \vec{R}\begin{pmatrix} x^0 & x^1 & x^2 & x^3 \\ \tau = ct & x & y & z \end{pmatrix} \quad \text{(b)}. \tag{4.7}$$

The usage of superscripts in Eq. (4.7b) is not accidental. When real values of coordinates are used, the difference between covariant and contravariant vector components needs to be emphasized (see Appendix 1, § 8), and, consequently, superscripts are used with contravariant components. So the square of the interval between events will be written in the form

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 = \sum_{i=1}^{4} dx_i^2 = g_{ik}\, dx_i\, dx_k; \quad \text{(a)}$$

Differing from zero:

$$g_{11} = 1, \quad g_{22} = 1, \quad g_{33} = 1, \quad g_{44} = 1.$$

$$ds^2 = dx^{0^2} - dx^{1^2} - dx^{2^2} - dx^{3^2} = g_{ik}\, dx^i\, dx^k. \quad \text{(b)} \qquad (4.8)$$

Differing from zero:

$$g_{00} = 1, \quad g_{11} = -1, \quad g_{22} = -1, \quad g_{33} = -1.$$

Note that the intervals given in Eqs. (4.8a) and (4.8b) have opposite signs. Since ds^2 may be either negative or positive, the choice of signs for ds^2 is of no practical importance.

The Lorentz transformation is a transformation of four-dimensional radius vector components, that is coordinates of an event. We shall write them out again:

$$x_1' = \Gamma (x_1 + iBx_4), \quad x_2' = x_2, \quad x_3' = x_3, \quad x_4' = \Gamma (x_4 - iBx_1); \quad \text{(a)}$$

$$x^{0'} = \Gamma (x^0 - Bx^1), \quad x^{1'} = \Gamma (x^1 - Bx^0), \quad x^{2'} = x^2, \quad x^{3'} = x^3. \quad \text{(b)} \qquad (4.9)$$

A four-dimensional radius vector is one of four-dimensional vectors, so that if in the reference frame K the following four-dimensional vectors are specified

$$\vec{A}(A_1A_2A_3A_4) \quad | \quad \vec{A}(A^0A^1A^2A^3),$$

the components of the same vectors will be determined as follows in the frame K':

$$A_1' = \Gamma (A_1 + iBA_4), \quad A_2' = A_2, \quad A_3' = A_3, \quad A_4' = \Gamma (A_4 - iBA_1); \quad \text{(a)}$$

$$A^{0'} = \Gamma (A^0 - BA^1), \quad A^{1'} = \Gamma (A^1 - BA^0), \quad A^{2'} = A^2, \quad A^{3'} = A^3. \quad \text{(b)} \qquad (4.10)$$

Eqs. (4.8a, b) represent the square of an infinitesimal vector $(d\vec{R})^2$. Consequently, the square of the norm of a four-dimensional

vector (which is an invariant quantity) must be determined in this way:

$$\vec{A}^2 = g_{ik}A_iA_k = = A_1^2 + A_2^2 + A_3^2 + A_4^2; \quad (a) \qquad \Big| \qquad \vec{A}^2 = g_{ik}A^iA^k = A_kA^k = = A^{0^2} - A^{1^2} - A^{2^2} - A^{3^2}. \quad (b) \tag{4.11}$$

Of course, Eqs. (4.11a) and (4.11b) give opposite signs for the invariant quantity $\vec{A}^2$. But this is of no significance just as in the case when the sign of the interval is determined (see the note after Eq. (4.8)). It should be borne in mind, though, that different signs of the interval alter the conditions defining "time-like" and "space-like" intervals and vectors. (There is no harmony in the literature concerning this issue.)

In Eq. (4.11b) we introduced covariant coordinates according to the formulae of Appendix 1, § 8: $A_k = g_{ik}A^i$. It is easy to notice that $A_0 = A^0$, $A_1 = -A^1$, $A_2 = -A^2$ and $A_3 = -A^3$.

Just as in the case of a three-dimensional space, we shall have to deal with tensors. Most easily the tensor component transformation law is derived from the transformation law for a product of two four-dimensional vector components. The transformation equations for the components of the four-dimensional vectors $\vec{A}$ and $\vec{B}$ can be written down using the symmetric notation (see Eqs. (2.40a, b)):

$$A_i = \tilde{a}_{il}A'_l, \quad B_k = \tilde{a}_{km}B'_m; \quad (a) \qquad \Big| \qquad A^i = a_{il}A'^l, \quad B^k = a_{km}B'^m. \quad (b) \tag{4.12}$$

Multiplying the left-hand and right-hand sides of these equations, we obtain at once the transformation rules for vector component products:

$$A_iB_k = \tilde{a}_{il}\tilde{a}_{km}A'_lB'_m, (a) \quad \Big| \quad A^iB^k = a_{il}a_{km}A'^lB'^m. \quad (b) \tag{4.13}$$

Thus we obtain the general transformation law for the tensors $T_{ik} = A_iB_k$ and $T^{ik} = A^iB^k$:

$$T_{ik} = \tilde{a}_{il}\tilde{a}_{km}T'_{lm}, \tag{4.14}$$

$$T^{ik} = a_{il}a_{km}T'^{lm}. \tag{4.15}$$

Eq. (4.15) in which the difference between the covariant and contravariant coordinates is essential, represents the transformation law for a twice-contravariant tensor.

In the 4-space the measured physical quantities should be so arranged as to possess quite definite transformation properties with respect to a transition from one IFR to another, i.e. to the Lorentz transformation. But in the coordinate transformation (in-

cluding the fourth coordinate of the Minkowski world) only tensor quantities possess the definite transformation properties, and tensors of different rank transform according to different rules. Hence, all physical quantities to which we ascribe a real meaning have to be tensors: either scalars, i.e. zero-rank tensors, or 4-vectors, i.e. first-rank tensors, or, finally, tensors of a higher-than-one rank. We shall see later that an electromagnetic field forms a second-rank tensor (see Chapter 6). The transition from customary three-dimensional quantities to four-dimensional ones (which is, no doubt, necessary in the case of the Lorentz transformation) is not always straightforward and is realized differently in different cases. It is often possible to represent, with some modification, a customary three-dimensional vector as a spatial part of a 4-vector. As to the fourth component, its expression seems to be rather surprising at first, but in the final analysis proves to be natural. There is nothing amazing in this since in a non-relativistic limit we nearly always come back from relativistic relationships to classical ones.

Chapters 5-7 provide numerous examples of constructing four-dimensional vectors and tensors.

§ **4.4. A pseudo-Euclidean plane.** Characteristic features of the pseudo-Euclidean space can be illustrated by means of the pseudo-Euclidean plane. One of the two coordinate axes must necessarily represent the time axis, or the axis of time-proportional quantities, since in the STR purely spatial geometry remains Euclidean, and only space-time is described by pseudo-Euclidean geometry. In our choice of reference frames it is most convenient to consider the plane (x, τ).

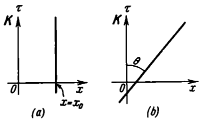


Fig. 4.1. (*a*) The world line of an object at rest at the point $x = x_0$. (*b*) The world line of an object moving uniformly along the x axis.

Recall that the four-dimensional space-time continuum whose points represent events is sometimes called the Minkowski world. Every event in our real physical world occurs at a definite world point of the Minkowski world. Considering a particle, one can regard its staying at a given point at a given moment of time as an event. No matter whether this particle moves or not, the sequence of events happening with the particle in the Minkowski world yields a certain curve called the *world line of the particle*.

Let us draw the x, τ axes of the frame K at right angles to each other and analyse the simplest cases. Let a particle be located at the point $x = x_0$ in the frame K; its world line in the

plane (x, τ) of the Minkowski world will be a straight line parallel to the τ axis (Fig. 4.1*a*). Let another particle move uniformly along the x axis in the frame K at the velocity v. Its world line in this frame will be a straight line inclined at the angle θ to the τ axis (Fig. 4.1*b*). A bit later we shall see that $\theta = \arctan (v/c)$.

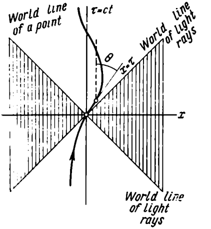


Fig. 4.2. The system of real coordinates x, $\tau = ct$. The particle's position at a given moment is specified by the point in this plane The particle's motion is depicted in this plane by the so-called world line of a point The world lines of motionless points are straight lines parallel to the τ axis The world line of light rays is the coordinate angle bisector. In the case of the variable velocity the angle formed by the tangent line to the world line and the τ axis is defined from the relation $\theta = \arctan (v/c)$, where v is the instantaneous velocity of a particle.

Now we shall examine an arbitrary motion of a particle in this reference frame. The motion of this particle is represented by the world line $x = x(\tau)$ in the plane (x, τ), as it is depicted in Fig. 4.2.

The inclination of the world line to the τ axis at each given point is determined by the derivative $dx/d\tau$ at that point. Indeed (see Fig. 4.2),

$$\tan \theta = \frac{dx}{d\tau} = \frac{1}{c} \frac{dx}{dt} = \frac{v}{c}. \quad (4.16)$$

Thus, the inclination angle is determined from the following equation:

$$\theta = \arctan \frac{v}{c} = \arctan \beta, \quad (4.17)$$

where $\beta = v/c$ and v is the instantaneous velocity of the point or the object. Inasmuch as $\beta < 1$ always, the angle θ cannot exceed 45° for any moving object. The world line of light rays will be represented by the bisecting line of the coordinate angle.

We saw in § 2.9 that the τ', x' axes are obtained from the τ, x axes as a result of the Lorentz transformation, provided these axes are drawn together in a scissors-like manner to the world line of light rays. The relativity of simultaneity is graphically seen in Fig. 4.3*a* where the τ', x' axes are drawn together with the τ, x axes. In the frame K' all events lying on the x' axis, or on the straight lines $\tau' = \text{const}$, are simultaneous. In terms of geometry, all these lines parallel to the x' axis represent the simultaneity lines in the frame K'.

Let us consider the two events A_1 and A_2 lying on the x' axis, both these events occurring simultaneously in the frame K' at the moment $t' = 0$. To find the moments of time at which these two events occur in the frame K, one should "project" these events on the τ axis by drawing straight lines parallel to the x axis,

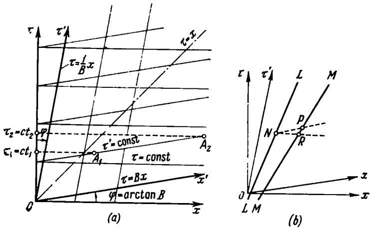


Fig. 4.3. (a) The Lorentz transformation reduces to the rotation of the x and τ axes through the angle $\varphi = \arctan B$ about the origin of coordinates toward the coordinate angle bisector and their new positions x', τ'. The straight lines $x' = \text{const}$ are now parallel to the $O\tau'$ axis, while the straight lines $\tau' = \text{const}$ are parallel to the Ox' axis (We have passed over to the rectilinear oblique-angled system of coordinates) The relativity of simultaneity is clearly seen: the events A_1 and A_2 which are simultaneous in the frame K' (lying on the straight line $\tau' = \text{const}$) are not simultaneous in the frame K To find the respective moments in the frame K, we project them on the τ axis by means of straight lines parallel to the x axis. (b) Here are two world lines of objects (LL and MM). The relativity of the distance between moving objects is seen very well To find the distance between them, one has to determine the coordinates of these objects simultaneously Let one of the objects be located at the point N. Then in terms of the frame K the second object is at the point R at the same moment. But in terms of the frame K' the second object is at the point P at the same moment. The sections NR and NP corresponding to the distances between the objects have different lengths.

since in the frame K' the events lying on the straight lines $\tau' = \text{const}$ (Fig. 4.3a) are simultaneous. We see that in the frame K these events occur at different moments of time t_1 and t_2. Of course, this is only a geometric illustration of the relativity of clock synchronization that we dealt with in § 2.4.

A very important result follows from Fig. 4.3b. It shows the world lines of two objects moving uniformly but at different velocities. To determine the distance between them at a given moment of time, the coordinates of these objects should be found simul-

taneously in the frame in which this distance is being determined. It is clearly seen that the distance between objects measured in the frames K and K' proves to be different. Due to the equivalence of reference frames none of the distances obtained can be regarded true. But then all laws of mechanics, in which force depends on distance, become ambiguous in the case of moving objects. Naturally, this problem did not emerge in Newtonian mechanics where time was regarded absolute.

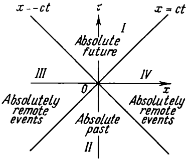


Fig. 4.4. The intersection of the space-time cone by the plane (x, τ). The point O represents event 1. All events located in quadrants III and IV represent absolutely remote events with respect to event O. The events located in quadrant I represent absolute future while the events located in quadrant II absolute past.

Let us consider the x, τ axes of the frame K (Fig. 4.4). The square of the interval between two world points is defined by the expression $s_{12}^2 = (\tau_2 - \tau_1)^2 - (x_2 - x_1)^2$. For the sake of simplicity let us suppose that event 1 occured at the point $x = 0$ at the moment $\tau = 0$, i.e. at the point O. Any events that occurred on the x axis before and after event 1 are depicted by points in the plane (x, τ). Since the square of the interval, that is the distance, from event 1 to any other event is equal to $s^2 = \tau^2 - x^2$, this plane is subdivided into four quadrants I, II, III, IV by the straight lines $x = \tau$, which correspond to the sequence of events consisting in the emission of a signal from the point $x = 0$ at the moment $\tau = 0$ and its arrival at the point x at the moment τ. The interval between the events located on the straight lines $\tau^2 - x^2 = 0$ is light-like, and the "distance" between such events is equal to zero in the pseudo-Euclidean plane. Now let us consider the four quadrants exterior to the light-like straight lines. In quadrant I $s^2 = \tau^2 - x^2 > 0$. Consequently, the interval between any event of quadrant I and event 1 is time-like. For all events of this quadrant $\tau > 0$; consequently, all of them will occur *after* event 1, and no choice of a reference frame can alter this situation. This means that quadrant I is the region of absolute future with respect to O. In quadrant II $s^2 > 0$ also, but here for all events $\tau < 0$; hence, quadrant II is the region of absolute past with respect to event 1.

In quadrants III and IV $s^2 < 0$, i.e. the interval between any event located in this region and event 1 is space-like. All these events occur at points which do not coincide with the point at

which event 1 occurred, and again it is impossible to alter this by the choice of a reference frame. However, one can find such reference frames where a given event from quadrant III or IV can happen before or after, or, finally, simultaneously with event 1, since the concepts "simultaneously", "before" and "later" are relative for the events located in this region.

If one examines two events located arbitrarily in the plane (τ, x), the character of the interval between them will be determined from the slope of the straight line connecting these two points. If the straight line is inclined to the x axis at the angle exceeding $\pi/4$, the interval between events 1 and 2 is time-like; if the angle is less than $\pi/4$, the interval is space-like. Finally, if this line is parallel to the bisecting line, the interval is light-like.

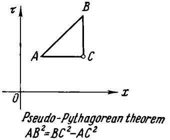


Fig. 4.5. The pseudo-Pythagorean theorem in the pseudo-Euclidean space.

In the four-dimensional space the equation describing the propagation of light has the form $c^2t^2 - x^2 - y^2 - z^2 = 0$. In terms of geometry this equation represents a "cone" in the four-dimensional space. Usually, this cone is called a *light cone*. The internal cavities of this cone correspond to the regions of "absolute future" and "absolute past". The light cone surface on which the light-like directions are located is remarkable owing to the fact that its position in the four-dimensional space remains invariable for every world point under all transitions from one IFR to another.

Let an event consist in the arrival of a light ray at a certain world point where an observer is located. Thus, we deal with the observation of light signals at a given point of space and at a given moment of time. The light rays can get at a given world point only along those directions of the four-dimensional space which lie on the "light cone of past" down to infinity (practically far enough in terms of light units). Each generatrix of this cone can be associated with the point on the spatial sphere of an infinitely great radius in whose centre the observer is located. Such an assumed sphere is used for the observation of celestial bodies and is called the sky sphere.

When depicting the pseudo-Euclidean plane on a sheet of paper, it should be remembered that we are used to such relations between the lengths of rulers, which are customary in the Euclidean plane. In Fig. 4.5 a right triangle is shown with the side AC equal to $x_2 - x_1$ and BC to $\tau_2 - \tau_1$. But in this plane $AB^2 = BC^2 - AC^2$, according to the definition of the square of the interval and con-

trary to the Pythagorean theorem; so this is the pseudo-Pythagorean theorem. Therefore, the comparison of lengths in the plane (x, τ) should be performed cautiously.

In the Euclidean plane (x, y) the locus of points equidistant from the origin of coordinates is defined by the equation of the circumference $r^2 = x^2 + y^2 = \text{const}$. In the pseudo-Euclidean plane (x, τ) where the square of the distance from the origin of coordinates is defined by the relationship $s^2 = \tau^2 - x^2$, the locus of points "equidistant" from the origin of coordinates (Fig. 4.6) will pattern four hyperbolas (s^2 is not necessarily positive). If one chooses the hyperbola for which $s^2 = 1$ and draws rays from the origin of coordinates till they intersect with this hyperbola, the section of each of such rays will determine the unitary "pseudo-Euclidean" length in the corresponding direction. It is possible to give the physical interpretation for plotting the hyperbola $s^2 = 1$. Let particles having various velocities but the identical lifetime $\tau_0 = 1$ be generated at the world point $\tau = 0$, $x = 0$. Then the locus of the world points at which these particles decay, will be the hyperbola $s^2 = 1$, and the world lines of these particles will be represented by the rays outgoing from the world point (0, 0) and reaching this hyperbola.

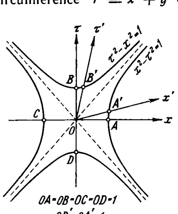


Fig. 4.6. The four equilateral hyperbolas $\tau^2 - x^2 = 1$, $x^2 - \tau^2 = 1$ are plotted in the coordinate system (x, τ). Since the Lorentz transformation leaves the expression $\tau^2 - x^2 = c^2t^2 - x^2$ invariant, we shall also obtain the hyperbolas $\tau'^2 - x'^2 = 1$, $x'^2 - \tau'^2 = -1$ in the new oblique-angled coordinate system. But this means that these four equilateral hyperbolas cross the axes x, τ, x', τ' at the distances from the origin equal to unity. The hyperbolas plotted are referred to as scale hyperbolas.

Let us consider the two pairs of equilateral hyperbolas in the plane (x, τ):

$$\tau^2 - x^2 = 1, \tag{4.18}$$

$$x^2 - \tau^2 = 1. \tag{4.19}$$

One can readily subdivide the plane (x, τ) into four quadrants, each containing one hyperbola. The dividing lines between the quadrants prove to be the asymptotes of these hyperbolas. Indeed, substituting the equation of the ray $\tau = kx$, passing through the origin of coordinates with the arbitrary slope k ($k = \tan\alpha$), into the equations of hyperbolas (4.18) and (4.19), we discover that

the intersection coordinate is determined from the equation $x^2 = \pm(1/(1-k^2))$. This equation has a real root only if $k^2 < 1$. When $k^2 = 1$, the coordinate of the intersection point on the x axis moves away into infinity. This means that the rays $\tau = x$ are asymptotes of these hyperbolas. Thus, the world lines of the light rays $x = ct$ are the asymptotes of the hyperbolas defined by Eqs. (4.18) and (4.19).

Each of these hyperbolas intersects only one of the axes: x or τ. The intersection points of the hyperbolas (4.19) with the x axis are determined from the condition $\tau = 0$. We see that the hyperbolas (4.19) intersect the x axis at the points $x = \pm 1$. In a similar way one can find that the hyperbolas (4.18) intersect the τ axis at the points $\tau = \pm 1$. Inasmuch as the hyperbolas (4.18) and (4.19) cut off the unitary sections on the coordinate axes, it is natural to call them the *scale hyperbolas*.

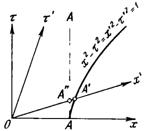


Fig. 4.7. The geometric illustration of the relativity of ruler lengths Quadrant I of Fig. 4.6 is depicted here. OA is a ruler at resl in the frame K. The world lines of its ends are $O\tau$ and AA''. The hyperbola $x^2 - \tau^2 = 1$ intersects the x axis at the point A and the x' axis at the point A'. Thus, $OA = 1$ and $OA' = 1$. To find simultaneously the position of the ruler's ends in the frame K', the world lines of the ruler's ends should intersect with some straight line $\tau' =$ const, for example, with the x' axis (corresponding to the moment $t' = 0$). Then the ruler's length in the frame K' turns out to be equal to OA''. But $OA'' < OA' = 1$.

Since the expression $\tau_2 - x^2 = c^2t^2 - x^2$ is the invariant of the Lorentz transformation, the equations $\tau'^2 - x'^2 = 1$, $\tau'^2 - x'^2 = -1$ will be valid in the frame K'. It follows directly that the same hyperbolas cut off the unitary sections on the new oblique-angled axes x' and τ' as well.

It is directly seen from Fig. 4.6 that the unitary sections of the x and x' axes are far from being equal. It should be remembered though that the representation of the pseudo-Euclidean plane in the Euclidean one is conditional and the "proper" units of length are identically chosen.

Now it becomes easy to explain in geometrical terms how the contraction of a moving ruler comes about. Let us show the x, τ axes and x', τ' axes in one figure, and plot that part of the hyperbola that passes through quadrant I of the coordinate systems K and K' (Fig. 4.7). The section OA represents a unitary ruler which is at rest in K. Its world lines in the frame K are straight lines parallel to the $O\tau$ axis and passing through the points O and A. But in terms of the frame K' the simultaneous position of the ends of the section OA at the moment $\tau' = 0$ corresponds to the intersection of its world lines with the x' axis, i.e. to the

points O and A''. The unitary ruler in K' is equal to OA'; it is seen from Fig. 4.7 that $OA'' < OA' = 1$.

Suppose now that a unitary ruler is at rest in the frame K' (Fig. 4.8). Then its length is equal to OA' and its world lines are parallel to the $O\tau'$ axis, one of them being the $O\tau'$ axis itself, and another the straight line $A'B$. In order to determine simultaneously the coordinates of the ruler's ends in terms of the frame K,

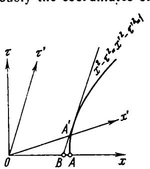


Fig. 4.8. The case is illustrated when a ruler is at rest in the frame K'. The world lines of its ends are straight lines parallel to $O\tau'$ (the $O\tau'$ axis itself and the straight line passing through B). The ruler's length in K is determined by the intersection of these world lines with the x axis ($t = 0$) and proves to be equal to OB But $OB < OA = 1$, and we obtain the same result: the ruler's length is the greatest in the frame where the ruler is at rest.

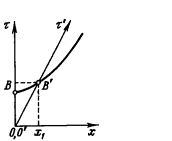


Fig. 4.9. The geometric illustration of the relativity of time intervals between two events Let a clock be at rest in the frame K' and be located at the origin of coordinates O. Its world line coincides with the $O\tau'$ axis. The reading of this clock at the world point B' differs by a unit from its reading at the point O'. But in the frame K the point B' is simultaneous with the world point B (lying on the same straight line $\tau = \text{const}$ with the point B') at which the clock (located at this point and at rest in the frame K) will indicate the time determined by the section OB relative to the reading of another clock from K located at the point O. It is seen from the figure that $OB < O'B' = 1$. This implies that the time interval, during which the clock from the frame K' moves, is less in terms of K' than in terms of K.

the world lines of the ruler's ends are to be intersected by any straight line $\tau = \text{const}$. It is more convenient for us to draw the straight line $\tau = 0$. From Fig. 4.8 it is seen that $OB < OA = 1$.

Let us dwell on a geometric illustration of the relativity of time intervals (Fig. 4.9). Let a clock be at rest at the origin of the coordinate system K'. Its world line will be the $O\tau'$ axis. At the moment of time $t = 0$ a moving clock was at the origin of the coordinate system K where we had its reading compared against one of the clocks of the system K' located at this point.

As before, we suppose that the clocks from both systems show the time $t = 0$ and $t' = 0$ at the moment when O and O' coin-

cide. Then the sections OB and $O'B'$ correspond to the time readings of the clocks of the systems K and K'.

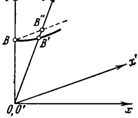


Fig. 4.10. The same as in the preceding figure, only now the clock is at rest at the origin of the frame K. The world line of the clock is the $O\tau$ axis. At the point B the clock will indicate a time unit The events lying on the straight line parallel to the x' axis and passing through the point B will be simultaneous with this moment in the frame K'. It is clear that $OB'' > OB' = 1$, i.e. a motionless clock will register the lesser time interval as compared to a moving clock.

At the world point B' the reading of the moving clock will increase by unity compared to that at the point O'. But the point B' in the frame K is simultaneous with all events located at the straight line $\tau = \text{const}$ passing through the point B'. In particular, the world line of the clock located at the point x_1 and at rest in K passes exactly through the point B'. This means that if the moving clock of K' registers the proper-time interval $O'B'$, the time interval registered by the two clocks of K (located at the points O and x_1) is equal to OB. It is seen in the figure that the time interval registered by the clock of K' is less, because $O'B' = 1$ and $OB > 1$.

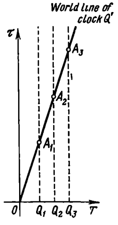


Fig. 4.11. The difference between the "proper" time of an object and the coordinate time registered by many clocks of the reference frame relative to which the object moves

And if the clock is at rest in the system K, it will register a time unit at the world point B (Fig. 4.10) which is simultaneous with the point B'' in the system K' (OB'' is the reading of the clock of the system K' which an observer from the system K will get at the point B''). The point B'' is obtained as a result of the intersection of the straight line parallel to the x' axis and passing through the point B with the $O\tau'$ axis. But $OB'' > OB' = 1$; consequently, the moving clock will again register the longer time interval than two motionless clocks. The length of the world line arc (in the pseudo-Euclidean plane!) is directly associated with the proper time of the object, being just proportional to it: $ds = c\,d\tau$. Hence, the length of the world line arc enables us to conjecture about the proper time that was registered by the clock fixed to the particle. It should be remembered, however, that one should be careful in the evaluation of the arc length in the pseudo-Euclidean plane. The "risk" is clearly visible from the fact that the

"arc length" for two points located at the finite spatial distance from each other may turn out to be equal to zero. Think for yourself why in the foregoing reasoning we obtained the correct results on the basis of geometry. Naturally, the peculiarities of the pseudo-Euclidean plane interfere with the interpretation of the results. As an example let us consider the difference between the proper time and the coordinate one, i.e. the time registered by the clock of the system relative to which an object moves. Let the clock Q' be at rest at the origin of the system K' and its world line be OA_3 (Fig. 4.11). As usual, the coinciding clocks at O and O' indicate $t = 0$, $t' = 0$.

The world lines of all clocks Q at rest in K are represented by straight lines parallel to the τ axis. At the world points A_1, A_2, $A_3, \ldots$ one can check the clock Q' against the clocks Q_1, Q_2, $Q_3, \ldots$ synchronized in K and indicating the common, unified for K, time at any world point A_1, A_2, $A_3, \ldots$. Its value at the world point A_1 is equal to the length of the world line Q_1A_1. For the clock Q', however, the length of the world line connecting O' and A_1 is equal to OA_1. But $OA_1^2 = Q_1A_1^2 - OQ_1^2$, from where it is clear that $OA_1 < Q_1A_1$. This implies that the clock Q' checked against the clocks Q_1, $Q_2, \ldots$ at rest in the frame K is slow compared to the clocks Q_1, $Q_2, \ldots$ synchronized in the frame K.

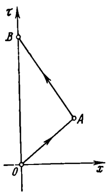


Fig. 4.12. The world lines of two "twins". The world line of the "traveller" is the broken line OAB, that of the "stay-at-home" the straight line OB The "traveller" undergoes an acceleration when he reverses his motion direction at the point B and thereby gets into a non-inertial reference frame for this time interval. The length of the world line of an object determines its proper-time interval. The proper-time interval is obviously less for the "traveller" than for the "stay-at-home" (see the pseudo-Pythagorean theorem in Fig. 4.5).

Finally, let two persons ("twins") be at the point O at first. Then one of them ("a traveller") moves uniformly and rectilinearly except for a short time interval needed to reverse the velocity direction before returning to the initial point O. The other "twin" remains at the point O all the time. It is seen from Fig. 4.12 that the world line of the "traveller" OAB is longer than that of the "stay-at-home". However, in accordance with the pseudo-Pythagorean theorem this means that the "traveller" spent less of his local time than the "stay-at-home" did. We shall come back again to this problem in Chapter 8.

CHAPTER 5

RELATIVISTIC MECHANICS OF A PARTICLE

The Einstein principle of relativity is valid provided that the basic laws of physics are formulated similarly throughout all inertial frames and differ only by the notation of variables associated with the given reference frame. In terms of physics the last statement implies that in every IFR measurements are carried out by means of instruments which are at rest in that frame. But the transformation of the coordinates of an event on transition from one IFR to another is the Lorentz transformation. Consequently, the equations of mechanics, for example, have to retain their appearance (in the above-mentioned sense) in any IFR. This condition is automatically fulfilled if the equations of mechanics are put down in a four-dimensional vector form. Indeed, in this case the transformation law of the left-hand and right-hand sides of such an equation is known, and it does not change the appearance of the equation. When put down in the vector (or the more general tensor) form, the equation is said to be written in the covariant form.

The Newtonian equation relating forces and accelerations is covariant relative to the Galilean transformation, although it is not covariant relative to the Lorentz transformation. However, the Lorentz transformation follows unambiguously from the Einstein postulates which are for certain confirmed experimentally. In order to satisfy the principal Einstein postulate on the equivalence of inertial frames of reference, one has to ensure the covariance of the equations of mechanics under the relativistic transformation of coordinates and time, i.e. the Lorentz transformation. The required equations of mechanics are fairly easy to write using the STR's four-dimensional geometric concept. We shall proceed in just this manner.

Certainly, the development of science does not cancel previously known ("correct") laws, but only sets limits to their application. There is always some conformity between various theories describing one and the same group of phenomena in extreme cases. The majority of equations of classical mechanics correspond to the extreme cases of relativistic equations with $\beta \to 0$. In other words,

classical mechanics is the extreme case of relativistic mechanics corresponding to the velocities which are small in comparison with that of light. Nevertheless, relativistic mechanics brings forward such conclusions that could not even be alluded to in the framework of classical mechanics (for example, the existence of the rest energy of an object).

§ 5.1. **A 4-velocity and 4-acceleration.** To write down the relations between physical quantities in space-time, we must construct the required 4-vectors. While doing this, we should remember that in the extreme case of small velocities the Lorentz transformation turns into the Galilean one, the relativity of time intervals and lengths does not manifest itself any more, and the Newtonian equations correspond to the Galilean principle of relativity provided it describes the transition from one IFR to another. In this extreme case time and space are not related, and we can utilize conventional three-dimensional quantities. Therefore, while composing four-dimensional quantities, we shall always try to make their three (spatial) components resemble the corresponding three-dimensional quantities. In the extreme case of small velocities $(\beta \to 0)$ the three components of four-dimensional quantities must turn into the conventional mechanical quantities.

We shall compose a 4-velocity and a 4-acceleration in the same way as we do the corresponding quantities in the three-dimensional space where a particle position is specified by the three-dimensional radius vector $\boldsymbol{r}$ and the 3-velocity is determined as the derivative of the radius vector with respect to time, $d\boldsymbol{r}/dt$. The 4-velocity cannot, however, be defined as a derivative of the 4-radius vector $\vec{R}$ with respect to time. To get the 4-vector velocity, we have to divide the 4-vector of the increment $d\vec{R}$ by a scalar (an invariant of the Lorentz transformation). But neither time nor its differential is a scalar.

One can take the interval or the proper time of a particle (see § 3.3) as an invariant time-dependent quantity. We shall introduce once more the proper-time concept, having associated it with the interval between events. We make use of the fact that the motion of a particle in the 3-space is a continuous sequence of events consisting in a particle occupying a definite point in space at a given moment of time. Let the coordinates of a particle in the frame K change by dx, dy, dz during the time dt, and its displacement be equal to $dl = \sqrt{dx^2 + dy^2 + dz^2}$. Consider the instantaneous inertial frame K' co-moving with the particle, i.e. the frame moving at the constant velocity V equal to the instantaneous velocity of the particle. In the frame K' the coordinates of the particle do not change during the infinitesimal time

interval dt': $dx' = dy' = dz' = 0$. The interval between events is invariant, so that

$$ds^2 = c^2\,dt^2 - dx^2 - dy^2 - dz^2 = c^2\,dt'^2.$$

In the frame K' the time interval dt' is the proper-time interval. In this chapter we shall designate it by $d\tau$ (we shall not use the designation $\tau = ct$ as we did in previous chapters). From the foregoing equation we have

$$d\tau = \frac{ds}{c} = \sqrt{1 - \frac{dx^2 + dy^2 + dz^2}{c^2\,dt^2}}\,dt =$$
$$= \sqrt{1 - \frac{1}{c^2}\left(\frac{dl}{dt}\right)^2}\,dt = \sqrt{1 - \frac{v^2}{c^2}}\,dt.$$

We have obtained the familiar result (§ 3.3) and demonstrated the invariance of the proper time ($d\tau = ds/c$). Here are the equations to be needed later:

$$d\tau = dt/\gamma, \quad ds = c\,d\tau, \quad \gamma = (1 - \beta^2)^{-1/2}, \quad \beta(t) = v(t)/c. \tag{5.1}$$

We see that the proper time of a particle is registered by a clock of an instantaneous co-moving IFR. But these instantaneous co-moving IFRs change during a finite time interval in the case of a particle moving with an acceleration. The final proper time of such a particle is defined as the overall time registered by many IFRs. As a matter of principle, the clock should not be rigidly linked with the particle, since any acceleration affects the clock rate. The proper time can be registered by the clock fixed rigidly to the particle only if the acceleration to which this particle is subjected does not affect the clock rate. The "proper time" can, however, be readily obtained from the time registered by the clock of the frame K (relative to which the particle moves) provided that the time dependence of the particle velocity, i.e. $v = v(t)$, is known:

$$\tau = \int \sqrt{1 - \frac{v^2}{c^2}}\,dt.$$

It is seen from the last equation and equations (5.1) that the coordinate time, that is the time registered by all clocks of K, is a function of the proper time τ. From the equation $ds = c\,d\tau$ one can see that in addition to the proper time $d\tau$ one may equally use the interval ds, with all equations differing by various powers of the invariant factor c.

Now let us introduce the 4-vector velocity

$$\vec{V} = \frac{d\vec{R}}{d\tau}. \tag{5.2}$$

Since $d\tau$ is an invariant and $d\vec{R}$ a vector, $\vec{V}$ is also a vector, no doubt. Let us disclose a three-dimensional connotation of the first three components of (5.2) in the notation of Eq. (4.7a):

$$u_a = \frac{dx_a}{d\tau} = \gamma \frac{dx_a}{dt} = \gamma v_a \quad (a = 1, 2, 3), \tag{5.3}$$

where v_a are the components of the conventional 3-velocity. Therefore, the first three components of the 4-velocity are those of the conventional 3-velocity multiplied by the factor γ depending on the absolute value of the particle velocity. The fourth component is to be found separately:

$$u_4 = \frac{dx_4}{d\tau} = \gamma \frac{d(ict)}{dt} = ic\gamma. \tag{5.4}$$

In accordance with the notation of Eq. (4.7) we have

$$u^0 = \frac{dx^{(0)}}{d\tau} = \gamma \frac{d(ct)}{dt} = \gamma c, \quad u^a = \frac{dx^a}{d\tau} = \gamma \frac{dx^a}{dt} = \gamma v_a.$$

Similarly to Eq. (4.7a, b) one can write

$$\vec{V}\begin{pmatrix} u_1 & u_2 & u_3 & u_4 \\ \gamma v_x & \gamma v_y & \gamma v_z & ic\gamma \end{pmatrix}; \ \text{(a)} \quad \Bigg| \quad \vec{V}\begin{pmatrix} u^0 & u^1 & u^2 & u^3 \\ \gamma c & \gamma v_x & \gamma v_y & \gamma v_z \end{pmatrix}. \ \text{(b)} \tag{5.5}$$

When $\beta \to 0$, i.e. when the velocity of an object $v \ll c$, the factor $\gamma \approx 1$, and the first three components of the 4-velocity of (5.5a) as well as the last three components of Eq. (5.5b) coincide with the conventional velocity. Of special interest is the fourth component of (5.5a) and the zeroth one of (5.5b) for the 4-velocity. They are different from zero even when a particle is at rest (if $v = 0$, $\gamma = 1$ and $u_4 = ic$, but $u^0 = c$). The last result has the obvious meaning: time cannot be stopped, it always flows without interruption. Accordingly, there is no quiescence in the four-dimensional world (in the sense that $\vec{V} \neq 0$). As to the "velocity of the time flow", it is defined by the choice of time units, of course.

The components of the 4-velocity can be also put down as follows:

$$\vec{V}(\gamma \mathbf{v}, ic\gamma); \ \text{(a)} \ \mid \ \vec{V}(c\gamma, \gamma \mathbf{v}). \ \text{(b)} \tag{5.6}$$

The square of the 4-vector is an invariant. It can be found from Eqs. (4.11a) and (4.11b) respectively:

$$\vec{V}^2 = \gamma^2 v^2 - c^2\gamma^2 = -c^2; \quad \vec{V}^2 = (c^2\gamma^2 - \gamma^2 v^2) = c^2.$$

The computation is easiest when made in the inherent reference frame of a particle at rest ($v = 0$). Then in (5.5a) only

$u_4 = ic$ will differ from zero, and in (5.5b) only $u^0 = c$. Consequently,

$$\vec{V}^2 = u_1^2 + u_2^2 + u_3^2 + u_4^2 = -c^2; \quad \text{(a)} \quad | \quad \vec{V}^2 = u^{0^2} - u^{1^2} - u^{2^2} - u^{3^2} = c^2; \quad \text{(b)} \qquad (5.7)$$

the squares of the 4-velocity in Eqs. (5.7a) and (5.7b) are opposite in sign due to the different determination of the interval (see Chapter 4). When the appropriate determination of $\vec{V}^2$ is chosen, however, this sign does not change, and it follows that $v < c$ in all cases.

As soon as the velocity in 4-space is written down in the form of a 4-vector, the transformation equations for the velocity components on transition from one inertial frame to another can be obtained at once. Let the components of the 4-velocity $\vec{V}$ (u_1, u_2, u_3, u_4) be specified in the frame K. In accordance with Eq. (4.10a) we shall obtain in the frame K'

$$u_1' = \Gamma(u_1 + iBu_4), \quad u_2' = u_2, \quad u_3' = u_3, \quad u_4' = \Gamma(u_4 - iBu_1), \qquad (5.8)$$

but the 4-velocities have the components $\vec{V}(\gamma v, ic\gamma)$, $\vec{V}'(\gamma' v', ic\gamma')$. Having substituted them in Eq. (5.8), we get

$$\gamma' v_x' = \Gamma(\gamma v_x - \gamma V), \quad \gamma' v_y' = \gamma v_y, \quad \gamma' v_z' = \gamma v_z,$$
$$ic\gamma' = \Gamma(ic\gamma - iB\gamma v_x). \qquad (5.9)$$

It follows from the last equation of (5.9) that

$$\frac{\gamma}{\gamma'} = \frac{1}{\Gamma\left(1 - \frac{V}{c^2} v_x\right)}. \qquad (5.10)$$

Substituting this expression in the first three equations (5.9),

$$v_x' = \frac{\gamma}{\gamma'} \Gamma(v_x - V), \quad v_y' = \frac{\gamma}{\gamma'} v_y, \quad v_z' = \frac{\gamma}{\gamma'} v_z,$$

we shall obtain the equations for the velocity components in K' which were derived in Chapter 3 from the Lorentz transformation.

Note, incidentally, that if in place of Eq. (5.8) of transition from K to K' one uses the equations for the reverse transition from K' to K, the following equation is obtained

$$\frac{\gamma}{\gamma'} = \Gamma\left(1 + \frac{V}{c^2} v_x'\right). \qquad (5.10')$$

instead of Eq. (5.10). This way we obtain the value of γ/γ' in terms of the velocity components in the frame K'. From Eq. (5.10'),

follows Eq. (3.41) derived otherwise here:

$$\sqrt{1-\beta^2} \equiv \frac{1}{\gamma} = \frac{1}{\gamma'\Gamma\left(1+\frac{V}{c^2}v'_x\right)} = \frac{\sqrt{1-\beta'^2}\sqrt{1-B^2}}{1+\frac{V}{c^2}v'_x}.$$

It follows from Eq. (5.10) that if a particle is at rest in K ($v=0$), then $\gamma' = \Gamma$; this result is obvious, because a particle which is at rest in K moves at the velocity $-V$ relative to K'.

The same result is obtained if one makes use of Eq. (5.6b) and the transformation equation (4.10b). We suggest that the reader do it himself. Our result is obvious: the spatial components of the 4-velocity determine the transformation of the conventional 3-velocity.

Now we are to define the 4-acceleration which we shall also construct as a 4-vector:

$$\vec{w} = \frac{d^2\vec{R}}{d\tau^2} = \frac{d\vec{V}}{d\tau}, \tag{5.11}$$

or expressed via components:

$$w_i = \frac{du_i}{d\tau} = \frac{d^2x_i}{d\tau^2}; \quad \text{(a)} \quad \Bigg| \quad w^i = \frac{du^i}{d\tau} = \frac{d^2x^i}{d\tau^2}. \quad \text{(b)} \tag{5.12}$$

Below we shall write out a few formulae dealing with acceleration, using the notation of Eq. (4.7a). They will be needed only in special cases. The four-dimensional acceleration components can be expressed by means of the three-dimensional components of the vectors $\boldsymbol{v}$ and $\dot{\boldsymbol{v}}$. We get

$$w_a = \frac{d}{dt}(\gamma v_a)\frac{dt}{d\tau} = \gamma v_a \frac{d\gamma}{dt} + \gamma^2\frac{dv_a}{dt} = \frac{v_a}{1-\beta^2} + \frac{v_a(\boldsymbol{v}\dot{\boldsymbol{v}})}{c^2(1-\beta^2)^2}, \tag{5.13}$$

because, as it is easy to verify,

$$\dot{\gamma} = \frac{d\gamma}{dt} = \gamma^3\beta\dot{\beta} = \frac{\gamma^3\boldsymbol{v}\dot{\boldsymbol{v}}}{c^2}, \tag{5.14}$$

and $dt/d\tau = \gamma$ The fourth component of the acceleration is

$$w_4 = \frac{d}{dt}(ic\gamma)\frac{dt}{d\tau} = ic\gamma\frac{d\gamma}{dt} = \frac{ic}{2}\frac{d\gamma^2}{dt} = ic\gamma^4\beta\dot{\beta} = \frac{i}{c}\frac{(\boldsymbol{v}\dot{\boldsymbol{v}})}{(1-\beta^2)^2}. \tag{5.15}$$

In the case of the uniform motion ($\dot{\boldsymbol{v}} = 0$) all four acceleration components turn into zero. In the reference frame in which a particle is at rest

$$w_1^0 = \dot{v}_x, \quad w_2^0 = \dot{v}_y, \quad w_3^0 = \dot{v}_z, \quad w_4^0 = 0, \tag{5.16}$$

i.e. the three spatial components of the 4-acceleration coincide with the conventional three-dimensional components of the accel-

eration, while the time component turns into zero. It is seen from Eq. (5.16) that

$$\vec{w}_0^2 = w_i^{02} = \dot{v}^2 > 0.$$

Due to the invariance of the square of the 4-vector norm (see Appendix I, § 1) one may regard the 4-vector acceleration as a space-like vector (see the definition of the interval (Eq. (4.5)).

Let us write out the components of the 4-vector acceleration $\vec{w}$ in the notation of Eq. (4.7):

$$\begin{aligned} &\vec{w}\left(\gamma \frac{d}{dt}(\gamma v),\ \frac{ic}{2}\frac{d\gamma^2}{dt}\right) \equiv \left(\gamma^2\dot{v} + \gamma^4\beta\dot{\beta}v,\ \frac{i\gamma}{mc}\frac{d\mathscr{E}}{dt}\right), \quad \text{(a)} \\ &\vec{w}\left(\gamma \frac{d}{dt}(c\gamma),\ \gamma\frac{d}{dt}(\gamma v)\right) \equiv \left(\frac{\gamma}{mc}\frac{d\mathscr{E}}{dt},\ \gamma\frac{d}{dt}(\gamma v)\right). \quad \text{(b)} \end{aligned} \tag{5.17}$$

The particle energy $\mathscr{E}$ introduced here will be defined later on (see Eq. (5.32) below). Using Eqs. (5.15) and (5.17), one can easily obtain

$$\vec{w}^2 = w_i^2 = \gamma^6\left(\dot{v}^2 - \frac{1}{c^2}[v\dot{v}]^2\right) > 0. \tag{5.18}$$

Now let us write out the transformation equation for the 3-acceleration ($\dot{v} = dv/dt$, $\dot{v}' = dv'/dt'$) on transition from one IFR to another. The Galilean transformation leaves the 3-acceleration of a particle invariable. The Lorentz transformation changes the 3-acceleration components. The simplest way to derive the transformation equation for the 3-acceleration components is as follows. Regarding v_x and v_x' as functions of t and t' respectively, and taking into account the relationship between t and t' (Eq. (2.16)), and, finally, designating $(v_x/c) = \beta_x$, $(v_x'/c) = \beta_x'$ etc., we shall obtain from Eq. (3.26)

$$dv_x = \frac{dv_x'}{\Gamma^2(1 + B\beta_x')^2}, \quad dv_y = \frac{dv_y' - B(\beta_y' dv_x' - \beta_x' dv_y')}{\Gamma(1 + B\beta_x')^2},$$

and in much the same way, dv_z. Having divided the left-hand and right-hand sides of these equations by the left-hand and right-hand sides of the equation $dt = \Gamma\left(dt' + \frac{B}{c}dx'\right)$, respectively, we get

$$\dot{v}_x = \frac{1}{\Gamma^3(1 + B\beta_x')}\dot{v}_x',$$

$$\dot{v}_y = \frac{\dot{v}_y' + B(\beta_x'\dot{v}_y' - \beta_y'\dot{v}_x')}{\Gamma^2(1 + B\beta_x')^3}, \quad \dot{v}_z = \frac{\dot{v}_z' + B(\beta_x'\dot{v}_z' - \beta_z'\dot{v}_x')}{\Gamma^2(1 + B\beta_x')^3}.$$

Of course, the same result will be obtained via the transformation of the 4-vector acceleration $\vec{w}$. In the frames K and K' it is

easiest to put it down in the form $\vec{w}(\gamma^2\dot{\boldsymbol{v}} + \gamma\boldsymbol{v}\dot{\gamma},\ ic\gamma\dot{\gamma})$, $\vec{w}'(\gamma'^2\dot{\boldsymbol{v}}' + \gamma'\boldsymbol{v}'\dot{\gamma}',\ ic\gamma'\dot{\gamma}')$.

From the transformation equations for the 4-vector components

$$w_1 = \Gamma(w_1' - iBw_4'), \quad w_4 = \Gamma(w_4' + iBw_1')$$

we get

$$\gamma\dot{v}_x + v_x\dot{\gamma} = \Gamma\frac{\gamma'}{\gamma}(\gamma'\dot{v}_x' + \dot{\gamma}'v_x' + V\dot{\gamma}'),$$

$$\dot{\gamma} = \Gamma\frac{\gamma'}{\gamma}[\dot{\gamma}' + B(\gamma'\dot{\beta}_x' + \dot{\gamma}'\beta_x')].$$

Using Eq. (5.10′), $\dot{v}_x$ will be found from these relations. Substituting $w_2 = w_2'$ and $w_3 = w_3'$ in the transformation equations, we obtain the equations for transformation of $\dot{v}_y$ and $\dot{v}_z$.

The transformation equation for the 3-acceleration components involves the velocity of a particle. But the 3-acceleration appears only when the velocity varies. Consequently, even when the 3-acceleration is constant in one IFR, it varies with time in all other frames: in relativistic mechanics a uniformly accelerated motion in one IFR is not such in all others.

§ 5.2. **A 4-force and a four-dimensional equation of motion.** Here are some three-dimensional classical relations to be referred to later on quite often. In classical mechanics a mass of a particle is treated as a constant. We shall designate it by m. The second law of Newton is put down as follows in classical mechanics:

$$\frac{d}{dt}(m\boldsymbol{v}) = \boldsymbol{F} \tag{5.19a}$$

or

$$\frac{d\boldsymbol{p}}{dt} = \boldsymbol{F}, \tag{5.19b}$$

where $\boldsymbol{F}$ is a three-dimensional vector of a conventional force; the quantity $\boldsymbol{p} = m\boldsymbol{v}$ is called a *classical momentum of a particle.*

Multiplying the left-hand and right-hand sides of Eq. (5.19a) by $\boldsymbol{v}\,dt$, we derive by means of simple transformations the corollary of the second law of Newton:

$$d(mv^2/2) = \boldsymbol{F}\boldsymbol{v}\,dt. \tag{5.20}$$

The right-hand side of Eq. (5.20) represents the work accomplished by the force $\boldsymbol{F}$; in accordance with the energy conservation law the left-hand side must incorporate the change of energy. Hence, the energy of a particle can be defined as $T = mv^2/2$ with an accuracy of a constant addendum. Here the addendum is adopted to be equal to zero, so that the particle at rest possesses no energy. Consequently, the energy $mv^2/2$ is associated only with

the motion of the particle; accordingly, it is appropriately called the *kinetic energy* (the "motion energy"). It should be pointed out that if we assumed that a motionless object possesses the energy $\mathscr{E}_0$ the "total" energy of a moving object would be $\mathscr{E} = T + \mathscr{E}_0$. The constant $\mathscr{E}_0$ can be interpreted as a permanent potential, or internal, energy. But in classical mechanics there is no reason to do this, and $\mathscr{E}_0$ can be treated as an arbitrary constant. Consequently, in classical mechanics the "total" energy of a free object may have either sign in principle (depending on the sign of the constant $\mathscr{E}_0$). Customarily $\mathscr{E}_0 = 0$, and the total energy of a free object coincides with the kinetic one.

Suppose now that a particle is in a potential field, i.e. the force acting on a particle can be expressed as $\boldsymbol{F} = -\operatorname{grad} U$, where $U(x, y, z)$ is a potential energy. Since $\boldsymbol{v}\,dt = d\boldsymbol{r}$ and $\operatorname{grad} U\,d\boldsymbol{r} = dU$, Eq. (5.20) will take the form

$$d\,(mv^2/2) = -\,dU.$$

from where follows the important law of classical mechanics, that is *the law of conservation of the total energy*:

$$\frac{d}{dt}(T + U) = 0;$$

in other words, $T + U = \text{const}$.

Now we can pass over to the definition of a 4-momentum of a particle $\vec{P}$. As in the case of the 3-momentum ($\boldsymbol{p} = m\boldsymbol{v}$) we specify the 4-momentum as the product of the invariant (scalar) mass m by the 4-velocity $\vec{V}$, so that $\vec{P} = m\vec{V}$. Therefore

$$\vec{P}(m\gamma\boldsymbol{v},\ im\gamma c);\quad \text{(a)}\quad \mid\quad \vec{P}(m\gamma c,\ m\gamma\boldsymbol{v}).\quad \text{(b)} \tag{5.21}$$

As it will be clear later, the invariant mass m is expediently called a *rest mass*. Analogously with Eq. (5.19) one may suppose that the four-dimensional equation of motion has the form

$$\frac{d\vec{P}}{d\tau} = \vec{F} \tag{5.22}$$

or in components

$$m\frac{du_i}{d\tau} = \mathfrak{F}_i, \tag{5.23}$$

where the differentiation is naturally accomplished with respect to the invariant proper time $d\tau$ (otherwise a vector relation will not be obtained), while the right part of the equation contains the 4-vector force $\vec{F}(\mathfrak{F}_1, \mathfrak{F}_2, \mathfrak{F}_3, \mathfrak{F}_4)$, whose components are still to be determined.

We shall recall once more (see § 4.3) why it is important to have the motion equation in a four-dimensional vector form (Eq. (5.22)). The point is that in accordance with the first postulate of Einstein all basic laws of physics must have the same form in all IFRs. In mathematical terms this means that equations describing physical laws have to be represented in the covariant form with respect to the Lorentz transformation. Equations are said to be written down in the covariant form if their left-hand and right-hand sides change alike under the Lorentz transformation. But this implies that the left-hand and right-hand sides must be correspondingly either scalar (invariant) quantities or 4-vectors, or tensors of the same rank (to be dealt with in Chapter 6). This is enough to ensure the invariance of relations presented in this form on transition from one IFR to another. Having put down the motion equations in the vector form (Eq. (5.22)), we ensured the covariance of this equation under the Lorentz transformation, i.e. the universal character of the Einstein principle of relativity.

The components of the vector $d\vec{P}/d\tau$ are familiar to us, for we know the components of $d\vec{V}/d\tau$ from Eq. (5.17a, b) and m is the invariant:

$$\begin{aligned} &\frac{d\vec{P}}{d\tau}\left(\gamma \frac{d}{dt}(m\gamma \boldsymbol{v}),\ i\gamma \frac{d}{dt}(m\gamma c)\right), \quad \text{(a)} \\ &\frac{d\vec{P}}{d\tau}\left(\gamma \frac{d}{dt}(m\gamma c),\ \gamma \frac{d}{dt}(m\gamma \boldsymbol{v})\right). \quad \text{(b)} \end{aligned} \tag{5.24}$$

We have denoted the components of the 4-vector $\vec{F}$ by the gothic letter $\mathfrak{F}$ supplied by indices, i.e. $\vec{F}(\mathfrak{F}_1, \mathfrak{F}_2, \mathfrak{F}_3, \mathfrak{F}_4)$ or $\vec{F}(\mathfrak{F}^0, \mathfrak{F}^1, \mathfrak{F}^2, \mathfrak{F}^3)$. Equating 4-vectors, we equate their components. The first three components of Eq. (5.24a) and the last three components of Eq. (5.24b) are obtained as follows ($\alpha = 1, 2, 3$):

$$\gamma \frac{d}{dt}(m\gamma v_\alpha) = \mathfrak{F}_\alpha; \quad \text{(a)} \quad | \quad \gamma \frac{d}{dt}(m\gamma v_\alpha) = \mathfrak{F}^\alpha. \quad \text{(b)} \tag{5.25}$$

Now let us determine the first three components of the 4-force $\mathfrak{F}_\alpha$. Obviously, they are proportional to the 3-force components since in the limit transition $\beta \to 0$ we have to get back to the conventional equation of Newton. If one retains the conventional definition of force and supposes, as before, that a "force determines a change of momentum", one should write

$$\mathfrak{F}_\alpha = \gamma F_\alpha, \quad \mathfrak{F}^\alpha = \gamma F_\alpha,$$

where F_α are the components of a conventional three-dimensional force. Having substituted the expressions for $\mathfrak{F}_\alpha$ into the right-

hand side of Eq. (5.25a), we obtain

$$\frac{d}{dt}(m\gamma v_\alpha) = F_\alpha \quad (\alpha = 1, 2, 3),$$

and having multiplied each of these equations by a corresponding unity coordinate vector $\boldsymbol{m}_\alpha$ and the relations thus obtained summed, we get the motion equation in a vector form:

$$\frac{d}{dt}(m\gamma \boldsymbol{v}) = \boldsymbol{F}. \tag{5.26}$$

Comparing Eq. (5.26) with the non-relativistic motion equation (5.19), we notice that they differ only in the definition of a momentum. The relativistic (three-dimensional) momentum is represented by the quantity

$$\boldsymbol{p} = m\gamma \boldsymbol{v}; \tag{5.27}$$

in this case Eq. (5.26) resembles Eq. (5.19) in outer appearance.

Thus, the three spatial components of Eq. (5.23) have yielded the second law of Newton in a relativistic form. However, the meaning of the fourth (or zeroth) relation is yet to be cleared up. In order to do this, $\mathfrak{F}_4$ (or $\mathfrak{F}^0$) should be known. But it turns out that having the three components of the 4-force determined, we thereby get the fourth component also determined. One can make sure of this in the following way. Differentiating Eq. (5.7a, b) with respect to τ, we get

$$\begin{aligned} &u_1\frac{du_1}{d\tau} + u_2\frac{du_2}{d\tau} + u_3\frac{du_3}{d\tau} + u_4\frac{du_4}{d\tau} = 0, \quad \text{(a)}\\ &u^0\frac{du^0}{d\tau} - u^1\frac{du^1}{d\tau} - u^2\frac{du^2}{d\tau} - u^3\frac{du^3}{d\tau} = 0. \quad \text{(b)}\end{aligned} \tag{5.28}$$

But according to Eq. (5.23) $du_i/d\tau = \mathfrak{F}_i/m$, with the first three components of $\vec{\mathfrak{F}}$ being determined by Eq. (5.26) and the components u_i by Eq. (5.5). Hence, Eq. (5.28a), for example, can be rewritten in the following form:

$$\gamma v_x = \frac{\gamma F_x}{m} + \gamma v_y \frac{\gamma F_y}{m} + \gamma v_z \frac{\gamma F_z}{m} + ic\gamma \frac{\mathfrak{F}_4}{m} = 0,$$

whence $\mathfrak{F}_4$ (and similarly $\mathfrak{F}^0$) can be derived at once:

$$\mathfrak{F}_4 = \frac{i\gamma}{c}(\boldsymbol{F}\boldsymbol{v}); \quad \text{(a)} \quad | \quad \mathfrak{F}^0 = \frac{\gamma}{c}(\boldsymbol{F}\boldsymbol{v}). \quad \text{(b)} \tag{5.29}$$

We give the reader a chance to derive Eq. (5.29b) by himself.

Thus, we have found the components of the 4-force $\vec{F}$ referred to as a *Minkowski force*:

$$\vec{F}\begin{pmatrix} \mathfrak{F}_1 & \mathfrak{F}_2 & \mathfrak{F}_3 & \mathfrak{F}_4 \\ \gamma F_x & \gamma F_y & \gamma F_z & i\frac{\gamma}{c}(\boldsymbol{F}\boldsymbol{v}) \end{pmatrix} \equiv \left(\gamma \boldsymbol{F},\ i\frac{\gamma}{c}(\boldsymbol{F}\boldsymbol{v})\right), \quad \text{(a)}$$

$$\vec{F}\begin{pmatrix} \mathfrak{F}^0 & \mathfrak{F}^1 & \mathfrak{F}^2 & \mathfrak{F}^3 \\ \frac{\gamma}{c}(\boldsymbol{F}\boldsymbol{v}) & \gamma F_x & \gamma F_y & \gamma F_z \end{pmatrix} \equiv \left(\frac{\gamma}{c}(\boldsymbol{F}\boldsymbol{v}),\ \gamma \boldsymbol{F}\right). \quad \text{(b)} \qquad (5.30)$$

To clear up the meaning of the fourth relation of Eq. (5.23) in terms of Eq. (5.5a), or the zeroth one in terms of Eq. (5.5b), the corresponding components in Eqs. (5.24) and (5.30) should be equated:

$$i\gamma \frac{d}{dt}(m\gamma c) = i\frac{\gamma}{c}(\boldsymbol{F}\boldsymbol{v}),$$

or otherwise

$$\frac{d}{dt}(m\gamma c^2) = \boldsymbol{F}\boldsymbol{v}. \qquad (5.31)$$

Here we can go over the same reasoning which was evolved in connection with Eq. (5.20). The right-hand side of Eq. (5.31) represents the work performed by the force; the left-hand side must contain the energy change. Let us define the total energy of a free relativistic particle as

$$\mathscr{E} = mc^2\gamma = mc^2(1-\beta^2)^{-1/2}; \qquad (5.32)$$

here $\beta = v/c$, where v is the absolute value of the three-dimensional velocity of the particle. It should be pointed out that the energy of the particle is determined from Eq. (5.31) with an accuracy of a constant value. Eq. (5.32) means that a motionless particle ($v = 0$, $\beta = 0$) possesses the energy $\mathscr{E}^0 = mc^2$. Such a value for the constant is not chosen at will, but comes from the limit transition to the classical velocity summation formula.

We shall postpone the discussion of the relativistic motion equation (5.26) and the relativistic energy relation (Eq. (5.32)) till §§ 5.3 and 5.4. In the meantime we shall dwell on the transformation of the 4-force and the consequences following from it. Let us write out the force transformation law (in terms of Eq. (5.30a)):

$$\mathfrak{F}_1' = \Gamma(\mathfrak{F}_1 + iB\mathfrak{F}_4), \quad \mathfrak{F}_2' = \mathfrak{F}_2, \quad \mathfrak{F}_3' = \mathfrak{F}_3, \quad \mathfrak{F}_4' = \Gamma(\mathfrak{F}_4 - iB\mathfrak{F}_1). \qquad (5.33)$$

We shall begin with a simple case. Let the three-dimensional force $\boldsymbol{F}$ act on a particle which is at rest in the frame K^0. Then in accordance with Eq. (5.30*a*) $\vec{F}^0(\boldsymbol{F}^0, 0)$. From Eq. (5.33) we ob-

tain

$$\gamma' F'_x = \Gamma F^0_x, \quad \gamma' F'_y = F^0_y, \quad \gamma' F'_z = F^0_z, \quad \gamma' \boldsymbol{F}' \boldsymbol{v}' = -\Gamma V F^0_x.$$

In the considered case a particle moves relative to K' at the velocity of the reference frame K^0, i.e. at the velocity $-V$. Consequently, $\gamma' = \Gamma$, and we obtain the component transformation formula for the force and the work accomplished by this force:

$$F'_x = F^0_x, \; F'_y = F^0_y \sqrt{1 - B^2}, \; F'_z = F^0_z \sqrt{1 - B^2}, \; \boldsymbol{F}' \boldsymbol{v}' = -VF^0_x. \quad (5.34)$$

It is seen from the first three relations (5.34) that the force components parallel to the relative motion velocity remain invariable. The force components normal to the relative motion velocity change. It is easy to find out the meaning of the last relation of (5.34). If the particle was at rest in the frame K^0, it moves at the velocity $-V$ in the frame K'. The work is performed only by the force component F_x (all other components being normal to the motion direction). The power developed by the force F^0_x in the frame K^0 is equal to $-F^0_x V$ which corresponds to the result that we obtained.

From Eqs. (5.34) one sees that in non-relativistic case, when $B \ll 1$, the three-dimensional force does not change on transition from one IFR to another. This fact wholly agrees with our intuitive ideas of the force invariance in any reference frame. However, in the presentation of the STR and, in particular, in the derivation of certain relations of the STR, when making use of the transformation of forces, one has to emphasize first of all the variation of force components on transition from one IFR to another.

In a general case, using Eq. (5.30a), we obtain from Eq. (5.33)

$$\gamma' F'_x = \Gamma \left[\gamma F_x - \frac{B}{c} \gamma (\boldsymbol{F}\boldsymbol{v}) \right], \quad \gamma' F'_y = \gamma F_y, \quad \gamma' F'_z = \gamma F_z,$$
$$\gamma' (\boldsymbol{F}'\boldsymbol{v}') = \Gamma [\gamma (\boldsymbol{F}\boldsymbol{v}) - V \gamma F_x].$$

Rewriting the last equations in the form

$$F'_x = \frac{\gamma}{\gamma'} \Gamma \left[F_x - \frac{B}{c} (\boldsymbol{F}\boldsymbol{v}) \right], \quad F'_y = \frac{\gamma}{\gamma'} F_y, \quad F'_z = \frac{\gamma}{\gamma'} F_z,$$
$$(\boldsymbol{F}'\boldsymbol{v}') = \frac{\gamma}{\gamma'} \Gamma [\boldsymbol{F}\boldsymbol{v} - V F_x] \quad (5.35)$$

and taking into account Eq. (5.10), we obtain finally

$$F'_x = \frac{F_x - \frac{B}{c} (\boldsymbol{F}\boldsymbol{v})}{1 - \frac{V}{c^2} v_x}, \quad F'_y = \frac{F_y \sqrt{1 - B^2}}{1 - \frac{V}{c^2} v_x},$$
$$F'_z = \frac{F_z \sqrt{1 - B^2}}{1 - \frac{V}{c^2} v_x}, \quad \boldsymbol{F}'\boldsymbol{v}' = \frac{(\boldsymbol{F}\boldsymbol{v}) - V F_x}{1 - \frac{V}{c^2} v_x}. \quad (5.36)$$

It is seen from the transformation formulae (Eq. (5.36)) for a 4-force that if there is no three-dimensional force in some IFR, such a force cannot appear in any other IFR. Thus, forces transform on transition from one IFR to another, but they never appear or disappear.

A validity of the law of inertia in all IFRs follows from this immediately. If in one IFR no forces act on an object and it moves due to inertia ($\boldsymbol{v} = \text{const}$), the same situation will be observed in any other IFR (see Eqs. (5.27) and (5.32)).

§ 5.3. A three-dimensional relativistic equation of motion of a particle (the second law of Newton in a relativistic form). Having written down the motion equation in a 4-vector form (Eq. (5.23)) and determined the components of the 4-force (the Minkowski force), we satisfied the principle of relativity for one thing, and, for another, obtained the four components of the motion equation. The three components provided us with the "motion equation" *per se* in a three-dimensional form (Eq. (5.27)), while the fourth component permitted us to determine the relativistic expression for energy (Eq. (5.32)). Eq. (5.27) was derived on the assumption that equations of dynamics must retain their appearance in all IFRs, i.e. they must be covariant with respect to the Lorentz transformation. However, even without passing from one IFR to another, we are aware that the exact equation of motion is represented by Eq. (5.26) and not by Eq. (5.19). Let us write out these two equations side by side and clarify the difference between them:

$$\frac{d}{dt}(m\boldsymbol{v}) = \boldsymbol{F}; \quad \text{(a)} \quad \Big| \quad \frac{d}{dt}(m\gamma\boldsymbol{v}) = \boldsymbol{F}. \quad \text{(b)} \qquad (5.37)$$

First of all, it is clear that when $\beta = v/c \ll 1$, i.e. $\gamma \approx 1$, Eq. (5.37*b*) passes into Eq. (5.37a). This means that classical mechanics is the extreme case of relativistic mechanics when particles move at non-relativistic velocities. Moreover, to satisfy the Galilean principle of relativity in classical mechanics, the Galilean transformation has to be valid, requiring $B \ll 1$ (see § 2.7), i.e. the relative velocity of the considered reference frames must be non-relativistic as well.

Sometimes, from the comparison of Eqs. (5.37a) and (5.37b) one may conclude that the difference between them consists in the fact that in Eq. (5.37b) mass depends on velocity, so that taking $m\gamma$ for a relativistic mass, we obtain a classical equation. Now we shall see that the things are much more complicated, and in Supplement IV we shall discuss why there is no sense in introducing a dependence of mass on velocity.

In order to compare Eqs. (5.37a) and (5.37b), the left-hand side of (5.37b) should be rewritten using Eq. (5.32) and the fol-

lowing identity (see also Eq. (5.31)):

$$\frac{d}{dt}(m\gamma v)=\frac{d}{dt}\left(\frac{\mathscr{E}}{c^2}v\right)=\frac{v}{c^2}\frac{d\mathscr{E}}{dt}+\frac{\mathscr{E}}{c^2}\frac{dv}{dt}=\frac{v}{c^2}(Fv)+m\gamma\frac{dv}{dt}.$$

Rearranging the terms, one may rewrite Eqs. (5.37a) and (5.37b) as follows:

$$m\frac{dv}{dt}=F, \quad \text{(a)}$$

$$m\frac{dv}{dt}=\frac{1}{\gamma}\left[F-\frac{v}{c^2}(Fv)\right]=\frac{1}{\gamma}[F-\beta(F\beta)]. \quad \text{(b)} \qquad (5.38)$$

It is noteworthy that in an IFR co-moving with a particle ($v=0$) Eq. (5.38a) coincides with Eq. (5.38b). It is immediately seen from these relations that the principal difference of the relativistic law of dynamics, Eq. (5.37b), from the classical one, Eq. (5.37a), lies in the fact that in the former the direction of a 3-acceleration does not coincide, generally speaking, with that of a force. Consequently, the straightforward comparison of the force and acceleration components, readily carried out for the case of Eq. (5.37a), is quite impossible here. It is seen from Eq. (5.38b) that there are still two cases for which an acceleration and a force are oriented along the same direction and the definitions of mass in Eqs. (5.37a) and (5.37b) can be directly intercompared.

(a) Let the force acting on a particle be always normal to its velocity, i.e. $F \perp v$. Then from Eq. (5.38b) the motion equation is immediately obtained in the following form:

$$m\gamma\frac{dv}{dt}=F, \qquad (5.39)$$

where, according to Eqs. (5.31) and (5.32), $\gamma=\text{const}$. This case is realized when a charged particle moves in a constant magnetic field. The Lorentz force $F=e\,[vB]$ is so directed that $Fv=0$ (always). It may be said that the motion of a relativistic particle in a constant magnetic field proceeds in accordance with the classical motion equation (5.19), but with some effective (but constant) mass $m\gamma$. This is valid, however, only for a special case, when the condition $Fv=0$ is satisfied. To make sure, let us consider another case.

(b) Let a force acting on a particle be always directed along its velocity. This implies, naturally, the rectilinear motion of a particle (in case of a definite choice of the initial velocity). A simple example of such a motion is provided by a charged particle moving in a plane capacitor when the initial velocity is directed along the electric field. If $F \parallel v$, then $v(Fv)=F(vv)=$

$= Fv^2$, and from Eq. (5.38b) we obtain the following motion equation:

$$m\gamma^3 \frac{dv}{dt} = F,$$

where γ is a variable quantity.

Thus, in the two special cases permitting of an intercomparison of Eqs. (5.38a) and (5.38b) we get a different dependence of mass on velocity; this indicates that there is no universal dependence of mass on velocity. It is sound practice to use the invariant rest mass (see Supplement IV).

As in classical mechanics, the equation of dynamics can also be written for the case when the rest mass of a particle varies due to the exchange of energy and momentum with the environment. If a particle loses the 4-vector momentum $\vec{\Pi}(\Pi_i) = \left(\gamma\Pi, \frac{i}{c}\gamma\Phi\right)$ per unit time due to convection, Eq. (5.22) should be replaced by

$$\frac{d\vec{P}}{d\tau} = \vec{F} + \vec{\Pi},$$

or

$$\frac{d}{d\tau} = (mu_i) = \mathfrak{F}_i + \Pi_i, \tag{5.40}$$

where only $\mathfrak{F}_i$ is a genuine mechanical force satisfying the condition $\vec{F}\vec{V} = 0$. Writing out the components of Eq. (5.40), we get

$$\frac{dP}{dt} = F + \Pi, \quad \frac{dE}{dt} = Fv + \Phi.$$

Here Π and Φ are the momentum and the energy delivered to a particle through convection per unit time. Having composed the product

$$\vec{\Pi}\vec{V} = \gamma^2(\Pi v - \Phi) = -\Phi^0,$$

we see that the quantity Φ^0 is the energy delivery rate in the frame in which the particle is at rest; it is equal to the rate of change of the particle's rest energy. Indeed, differentiating Eq. (5.40), we obtain

$$m\frac{du_i}{d\tau} + u_i\frac{dm}{d\tau} = \mathfrak{F}_i + \Pi_i. \tag{5.41}$$

Multiplying the left-hand and right-hand sides of Eq. (5.41) by u_i and taking into account that $\vec{V}(d\vec{V}/d\tau) = 0$ and $\vec{F}\vec{V} = 0$, we get

$$\Phi^0 \equiv -\Pi_i u_i = c^2\frac{dm}{d\tau} = \frac{dE_0}{d\tau}.$$

Assuming that the genuine mechanical force must satisfy the condition

$$m\frac{du_i}{d\tau}=\mathfrak{F}_i,$$

the following quantity should be treated as a mechanical force (see Eq. (5.41)):

$$m\frac{du_i}{d\tau}=\mathfrak{F}_i+\Pi_i-u_i\frac{dm}{d\tau}=\mathfrak{F}_i+\Pi_i-\frac{\Phi^0}{c^2}u_i=\mathfrak{F}_i+\Pi_i+u_i\frac{\Pi_k u_k}{c^2},$$

in the case of the convective transfer of momentum and energy. When the genuine mechanical force is absent, the mechanical "reactive" force should be taken into consideration:

$$R_i=\Pi_i+u_i\frac{\Pi_k u_k}{c^2}.$$

This force satisfies the condition $\vec{R}\vec{V}=0$.

In a particular case the momentum Π can be attained not by mechanical means but, for example, via radiation or heat exchange between the particle and the environment. In the case of a pure heat transfer the 4-momentum of heat transferred to the particle during the time $d\tau$ is equal to $\delta Q_i \equiv \Pi_i d\tau = (\delta \boldsymbol{p},\ i(\delta Q/c))$.

Consequently, $\delta \boldsymbol{p}=\boldsymbol{\Pi}dt$, $\delta Q=\Phi dt$ are the momentum and the energy of the heat transferred during the time dt. In this case the 4-momentum of the transferred heat is a 4-vector.

§ 5.4. **The relativistic expression for a particle's energy.** Note that Eq. (5.31) can be obtained not only as the fourth component of Eq. (5.23), but directly from Eq. (5.37b) and exactly in the same way as Eq. (5.20) follows from Eq. (5.19) in classical mechanics. To make sure, let us multiply scalarwise both the left-hand and right-hand sides of Eq. (5.38b) by $\boldsymbol{v}$. We obtain

$$m\boldsymbol{v}\dot{\boldsymbol{v}}=\frac{1}{\gamma}\left[\boldsymbol{F}\boldsymbol{v}-\frac{v^2}{c^2}(\boldsymbol{F}\boldsymbol{v})\right]\equiv\frac{1}{\gamma^3}\boldsymbol{F}\boldsymbol{v}.$$

This is just Eq. (5.31) with Eq. (5.14) taken into consideration. On the other hand, the fourth component of Eq. (5.23) turned out to yield the energy conservation law. The expression for energy given by Eq. (5.32) differs essentially from the classical one. In Newtonian mechanics the energy of a motionless free object (i.e. an object possessing no potential energy) is assumed to be zero, so that the kinetic energy is clearly defined as $T=mv^2/2$. In relativistic mechanics the total energy of a free particle is defined as $\mathscr{E}=mc^2\gamma$. We call this energy "total" because it includes the energy of a resting object (the rest energy mc^2). But we mentioned earlier that Eq. (5.31) defines the energy with an accuracy of a constant; consequently, having selected the appropriate value

for it (by assuming $\mathscr{E}_0 = -mc^2$), one could have regarded, as in Newtonian mechanics, the energy of a resting object equal to zero. But it is impossible to do so in the STR. In STR mechanics one should not forget about the transformation rules for various quantities and the principle of correspondence with classical mechanics: many classical and relativistic quantities must coincide in the extreme case $\beta \to 0$. The Lorentz transformation is known to turn into the Galilean transformation at small relative velocities of reference frames ($B \to 0$, $\Gamma \to 1$); at small velocities of particles ($\beta \to 0$) a three-dimensional relativistic momentum turns into a three-dimensional classical one, $m\gamma v \to mv$. Suppose we determined the total energy of a free relativistic particle in the form $\mathscr{E} = mc^2\gamma + C$; then in the extreme, case $\beta \to 0$ we should obtain $\mathscr{E} = mc^2 + C$. Let us examine now the transformation of the 4-momentum components on transition from one IFR to another. It is performed in accordance with the following equations:

$$P_1' = \Gamma (P_1 + iBP_4), \quad P_2' = P_2, \quad P_3' = P_3, \quad P_4' = \Gamma (P_4 - iBP_1). \quad (5.42)$$

Substituting the values of the 4-momentum components P_1, P_2, P_3, and P_4 from Eq. (5.21), we obtain

$$p_x' = \Gamma \left(p_x - \frac{B}{c}\mathscr{E}\right), \quad p_y' = p_y, \quad p_z' = p_z, \quad \mathscr{E}' = \Gamma (\mathscr{E} + Vp_x), \quad (5.43)$$

where p_x, p_y, p_z and p_x', p_y', p_z' are the components of the three-dimensional relativistic momentum $p = m\gamma v$. In the extreme case corresponding to a transition to classical mechanics when $B \to 0$, $\beta \to 0$ and $p_x' \to mv_x'$, $p_x \to mv_x$, $\mathscr{E} \to mc^2$, the first relation of (5.43) would lead to $mv_x' = mv_x - mV - \frac{VC}{c^2}$. But the latter equation must yield the classical law of velocity summation: $v_x' = v_x - V$. It will indeed be the case if $C = 0$. That is how we prove the validity of Eq. (5.32). It should be noted that the principle of correspondence between the classical and the relativistic expressions for energy is not valid only because in the framework of the Newtonian mechanics one could not detect the existence of the rest energy, and the additive constant was chosen without allowing for the rest energy (see below).

It is seen from Eq. (5.32) that the total energy does not turn into zero even when the velocity of an object is equal to zero ($\gamma = 1$ at $v = 0$). The energy of a free particle in the frame in which it is at rest is equal to mc^2 and is called the *rest energy* $\mathscr{E}_0$. Although we dealt with a particle until now, its elementary character was not discussed. Therefore, all equations derived are quite applicable to any complex object (system) composed of diverse components. Naturally, m will then represent the total mass of an object, and v the velocity of its motion as a whole. The

equation $\mathscr{E}_0 = mc^2$ is valid for any object that rests as a whole. Consequently, the object's rest mass defines the total energy confined in it regardless of the origin of this energy.

In classical mechanics the energy of a resting object may be either positive or negative: it is defined with an accuracy of a constant value. In relativistic mechanics the energy of a free object (or the energy of a closed system) is always positive and related to the object's rest mass; the rest mass of an object defines its rest energy. The inertia of an object proved to be a measure of the object's energy. Whenever the object's energy changes by $\Delta\mathscr{E}$, the mass of this objects varies by $\Delta m = \Delta\mathscr{E}/c^2$.

The question arises as to how such a high energy could remain unnoticed, that is the rest energy confined within an object. In fact, one gram of a substance contains about 10^{21} ergs. The total amount of energy confined within a system is not so essential, however, as the part of energy that can be utilized. Although any mass stores a huge amount of energy, its realization is far from being easy. Only recently people have learned to employ the atomic energy. Until a short while ago the rest energy had not been realized (and, accordingly, mass had always been constant). Since the realized energy always represents a difference of energies, the existence of the rest energy did not manifest itself in any way.

Since c^2 is very large, a change of mass accompanying a change of an object's energy is very small and cannot be experimentally detected even though weighting was always one of the most precise kinds of measurement. For example, 1 kg of water heated by 100 degrees gains only $5 \cdot 10^{-9}$ g of mass. Such a minute mass change cannot be detected even by means of a most sensitive modern balance. The formation of nuclei, however, involves quite perceptible mass changes; moreover, this mass defect determines the binding energy (see § 5.6).

In relativistic mechanics it is natural to define the part of the particle's energy that turns into zero at $v = 0$ as a kinetic energy. It is obtained by deduction of the rest energy from the total energy of a particle:

$$T = \mathscr{E} - mc^2 = mc^2(\gamma - 1). \tag{5.44}$$

The same result can be obtained when a force's work is calculated according to the equation of relativistic dynamics:

$$dT = \boldsymbol{F}\boldsymbol{v}\,dt = \boldsymbol{v}\,d(m\gamma\boldsymbol{v}) = m\gamma\boldsymbol{v}\,d\boldsymbol{v} + mv^2\,d\gamma =$$

$$= m\gamma\boldsymbol{v}\,d\boldsymbol{v} + mv\gamma^3\beta\,(\beta\,d\beta) = m\gamma^3 v\,dv\left(\frac{1}{\gamma^2} + \frac{v^2}{c^2}\right) = m\gamma^3 \boldsymbol{v}\,d\boldsymbol{v}.$$

Hence

$$T = m\int \gamma^3 v\,dv = mc^2\gamma + \text{const.}$$

If $T = 0$ at $v = 0$ (i.e. at $\gamma = 1$), const $= -mc^2$, whence it follows again that $T = mc^2(\gamma - 1)$.

Let us find the conditions under which Eq. (5.44) is converted into an expression for the classical kinetic energy. Expanding γ in series

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} = 1 + \frac{1}{2}\beta^2 + \frac{3}{8}\beta^4 + \ldots,$$

we see that

$$T = \frac{mv^2}{2} + m\frac{3}{8}\frac{v^4}{c^2} + \ldots, \tag{5.45}$$

i.e. the classical kinetic energy has the meaning inasmuch as $\beta \ll 1$ and the term β^4 can be ignored.

Having designated the classical kinetic energy by T_{cl} and the relativistic one by T_{rel}, Eq. (5.45) can be rewritten as follows:

$$T_{rel} = T_{cl} + \frac{3}{8}m\frac{v^4}{c^2} + \ldots.$$

The ratio T_{rel}/T_{cl} is given by the following expression:

$$\frac{T_{rel}}{T_{cl}} = 1 + \frac{3}{4}\beta^2 + \text{(terms of the order of } \beta^4 \text{ and higher).}$$

In nuclear physics, where "the limits of applicability of Newtonian mechanics" have to be defined more accurately, it is assumed that $T_{rel} = T_{cl}$ provided the second term on the right-hand side of the equation is less than one per cent (remember that $\beta \ll 1$, and the series descends rapidly). Consequently, the boundary limit (which is, of course, conditional) may be found from the equation $\frac{3}{4}\beta^2 = 0.01$, $\beta \approx 0.12$.

Since $\gamma = \gamma(v)$, and $\gamma = \mathscr{E}/mc^2$ for a particle, one may speak of relativistic velocities when the total energy $\mathscr{E}$ of a particle exceeds appreciably its rest mass, i.e. when the condition $(\mathscr{E}/mc^2) \gg \gg 1$ is met. Surely both conditions are qualitatively equivalent, but it should be borne in mind that when the velocity approaches its limit (c), the energy tends to infinity. Therefore, very small velocity variations in the vicinity of c may alter radically the particle's energy.

In nuclear physics it is more convenient to deal with energies of particles rather than with their velocities. Of course, particles of different masses will have different energy limits defining the applicability of classical mechanics. For example, in the case of electrons this limit is equal to 3 kev, while for protons 7 Mev (it would be useful for you to try to obtain these numbers yourself).

Eq. (5.32) may be rewritten as

$$\mathscr{E} = \gamma\mathscr{E}_0 \tag{5.46}$$

if the zero energy $\mathscr{E}_0 = mc^2$ is taken into account. The rest energy $\mathscr{E}_0$ comprises all kinds of energy possessed by an object (or a system). Eq. (5.46) shows that all kinds of energy increase γ-fold on transition from the proper (co-moving) frame to any other inertial frame. There was nothing of the kind in classical mechanics. On the other hand, the total energy of a particle (Eq. (5.32)) and its kinetic energy (Eq. (5.45)) grow without bound when $v \to c$. This result has a plain physical meaning. A particle whose rest mass differs from zero cannot attain the velocity equal to c. This can be inferred from the fact that the particle would require an infinite energy in order to achieve that velocity. Here the limitedness of the velocity of light *in vacuo* shows up again. When treating light quanta (photons) as relativistic particles (see § 7.6), one should bear in mind that they belong to another class of particles and could not come into being as a result of acceleration of conventional particles, that is through a dynamic transition. The limit transition $v \to c$ is carried out in nature, but the terminal point of this transition ($v = c$) is never reached.

§ 5.5. A 4-vector of energy-momentum. The fourth (or zeroth) component of the 4-momentum of a free particle has a direct bearing on the particle's energy. This is evident from the following straightforward transformation:

$$P_4 = mic\,\gamma = \frac{i}{c} mc^2\gamma = \frac{i}{c}\mathscr{E}; \quad \text{(a)} \quad \Big| \quad P^{(0)} = mc\gamma = \frac{\mathscr{E}}{c}. \quad \text{(b)}$$

Consequently, the 4-vector $\vec{P}$ is referred to as a *4-vector of energy-momentum of a particle*. From Eq. (5.7) and from the fact that $\vec{P} = m\vec{V}$ it follows that

$$\vec{P}^2 = -m^2c^2; \quad \text{(a)} \quad | \quad \vec{P}^2 = m^2c^2. \quad \text{(b)} \qquad (5.47)$$

The component transformation law for a 4-vector $\vec{P}$ is given by Eqs. (5.42) and (5.43). It remains to rewrite Eq. (5.22) in the final form

$$\vec{P}\left(\boldsymbol{p},\, i\frac{\mathscr{E}}{c}\right); \quad \text{(a)} \quad \Big| \quad \vec{P}\left(\frac{\mathscr{E}}{c},\, \boldsymbol{p}\right). \quad \text{(b)} \qquad (5.48)$$

where $\boldsymbol{p} = m\gamma\boldsymbol{v}$. Let us consider a particle in the reference frame in which its relativistic 3-momentum $\boldsymbol{p} = m\gamma\boldsymbol{v}$ is equal to zero. The frame in which a particle is at rest ($\boldsymbol{v} = 0$) may be called a *proper reference frame*. Let the particle's energy in this frame be equal to $\mathscr{E}^0$. Then in the frame K' in accordance with Eq. (5.43)

$$\mathscr{E}' = \Gamma\mathscr{E}^0, \quad p'_x = -\Gamma\frac{\mathrm{B}}{c}\mathscr{E}^0 = -\frac{\mathrm{B}}{c}\mathscr{E}' = -\frac{\mathscr{E}'}{c^2}V. \qquad (5.49)$$

It is seen from Eq. (5.49) that the energy transferred by a particle is associated with the origination of a momentum. Indeed, in the proper frame a particle possesses the energy $\mathscr{E}^0$ which does not move in space. The momentum of a particle (an energy carrier) in this frame is equal to zero. In the frame K' the particle moves; its velocity is equal to $-V$. This means that the energy "flows" at this velocity. Eq. (5.49) for p'_x shows that the energy flow involves the momentum $p'_x = -\Gamma \frac{\mathscr{E}^0}{c^2} V$. This momentum coincides with a relativistic three-dimensional momentum because according to Eq. (5.32) $\mathscr{E}^0/c^2 = m$; the particle's velocity is equal to $-V$, and Γ coincides with γ in this case.

Thus, an energy carrier (a particle, in this case) needs a momentum to be attributed to. Although we have obtained this result for a particle, it has a general significance; we shall come across it again when examining an electromagnetic field (Chapter 6).

We would like to emphasize here that the very fact of integrating certain quantities into a 4-vector points to an intimate connection between them. The quantities which are 4-vector components (usually a 3-vector and a scalar) constitute in a sense a closed combination: to calculate the energy and momentum of a particle in the frame K', one should know those in the frame K (see Eq. (5.43)). The fourth (or zeroth) component of a 4-vector of energy-momentum cannot turn into zero. If in some frame the energy and momentum of a particle turn into zero, they are equal to zero in any other reference frame. This is where relativistic and classical relations differ fundamentally. In classical mechanics the energy and momentum of a stationary particle are equal to zero.

The square of a 4-vector is an important invariant. Let us write it out:

$$\vec{P}^2 \equiv p^2 - \frac{\mathscr{E}^2}{c^2} = -m^2c^2; \quad \text{(a)} \quad \Big| \quad \vec{P}^2 \equiv \frac{\mathscr{E}^2}{c^2} - p^2 = m^2c^2, \quad \text{(b)} \quad (5.50)$$

(we have used Eqs. (5.47) and (5.48)). Needless to say that it is equal in magnitude but opposite in sign in these two cases. What is important, we have found an invariant relationship between a relativistic momentum and a relativistic energy of a particle; it is essential in this case that the invariant equation (5.50) determines an invariant mass, a rest mass of a particle.

From Eq. (5.50) one can express the particle's energy in terms of its momentum:

$$\mathscr{E} = c\sqrt{p^2 + m^2c^2}. \quad (5.51)$$

The particle's energy expressed in terms of its momentum is referred to as the Hamiltonian function $\mathscr{H}$ of a particle. Thus,

Eq. (5.51) gives the Hamiltonian function of a particle. It is common knowledge that the derivative of the Hamiltonian function with respect to the momentum components yields the velocity components of a particle:

$$\frac{\partial \mathscr{E}}{\partial p_x} \equiv \frac{\partial \mathscr{H}}{\partial p_x} = \frac{dx}{dt} = v_x, \ \ldots, \quad \text{or} \quad \boldsymbol{v} = \nabla_p \mathscr{E}. \tag{5.52}$$

Eq. (5.52) may be derived by differentiating Eq. (5.50):

$$p_x dp_x + p_y dp_y + p_z dp_z = \frac{1}{c^2} \mathscr{E}\, d\mathscr{E}, \tag{5.53}$$

whence

$$\frac{d\mathscr{E}}{dp_x} = p_x \frac{c^2}{\mathscr{E}}, \quad \frac{d\mathscr{E}}{dp_y} = p_y \frac{c^2}{\mathscr{E}}, \quad \frac{d\mathscr{E}}{dp_z} = p_z \frac{c^2}{\mathscr{E}}. \tag{5.54}$$

But inasmuch as $\mathscr{E} = mc^2\gamma$, and $\boldsymbol{p} = m\gamma\boldsymbol{v}$, then

$$\boldsymbol{p} = \frac{\mathscr{E}}{c^2} \boldsymbol{v} \tag{5.55}$$

for a particle, and we get back to Eq. (5.52).

In what follows we shall need a formula which is easy to derive from Eq. (5.52): multiplying the left-hand and right-hand sides of Eq. (5.52) by $d\boldsymbol{p}$, we obtain $\boldsymbol{v}d\boldsymbol{p} = d\mathscr{E}$. Those who are not particularly inclined toward a treatment of gradients should note that the same result follows at once from Eqs. (5.37a) and (5.37b) both in the classical and relativistic cases although a momentum is defined there in different ways, of course. Multiplying the left-hand and right-hand sides of Eqs. (5.37a) and (5.37b) by $\boldsymbol{v}\,dt$, we obtain $\boldsymbol{F}\,d\boldsymbol{r}$, i.e. $d\mathscr{E}$, in the right-hand side, and $v\,d\boldsymbol{p}$ in the left-hand one. Thus, in both classical and relativistic mechanics we have the same formula

$$d\mathscr{E} = \boldsymbol{v}\,d\boldsymbol{p}, \tag{5.56}$$

but the definition of energy and momentum will be different.

In the simplest case of a unidimensional motion $\mathscr{E} = \mathscr{E}(p)$ is determined according to Eq. (5.51). In the plane of the variables $\mathscr{E}$, p velocity is determined via a tangent of a slope angle of a curve $\mathscr{E} = \mathscr{E}(p)$ at a given point. When $p \gg mc$, Eq. (5.51) becomes

$$\mathscr{E} = cp. \tag{5.51$'$}$$

Particles with a finite rest mass are called *ultra-relativistic* if the last relation holds for them. Eq. (5.51′) is valid both for ultra-relativistic particles and for photons. We shall see in § 7.6 that light quanta (photons) may be treated as relativistic particles. But now it is worthwhile to point out that they are quite special particles. In any IFR these particles possess a finite momentum

and a finite energy. Their velocity *in vacuo* is the same in any IFR. They cannot originate from any of the particles possessing a finite rest mass by means of their acceleration. Finally, we infer from Eq. (5.50) that the photon's rest mass is equal to zero.

Let us consider some more relationships for a particle located in an external potential field. Since the field is assumed to be propagating at a finite velocity, the primary principle of the STR, the finiteness of the signal transmission velocity, is observed. As to the force acting on a particle, it is defined from the magnitude of the potential function at the point in which the particle is located (the field is stationary).

When a particle is located in a potential field, then $\boldsymbol{Fv}\,dt = -dU$ and Eq. (5.31) turns into $d(mc^2\gamma) = -dU$, whence follows the law of total energy conservation for a relativistic particle in a potential field:

$$mc^2\gamma + U = \text{const.} \tag{5.57}$$

(The energy is total in the sense that the sum of the relativistic energy and potential energy of a particle remains constant.) In relativistic mechanics the kinetic energy is equal to $mc^2(\gamma - 1)$; having altered the magnitude of the constant in the right-hand side by mc^2, the energy conservation law can be rewritten in the form

$$mc^2(\gamma - 1) + U = \text{const.} \tag{5.58}$$

When a particle is located in a conservative field, its velocity and potential energy may change in the course of motion but the value of $\mathscr{E} = mc^2\gamma + U$ remains constant (see Eq. (5.57)) in that it is time independent in a given IFR. The quantity $\mathscr{E}$ may be called *a total energy of a particle in a conservative field*. Surely, this quantity remains constant in any IFR but changes its (invariable) value for another on transition from one IFR to another. The definition of a 4-momentum for a particle in a conservative field as $\vec{P} = m\vec{V}$ holds good, but Eq. (5.48) is to be replaced by $\vec{P}\left[m\gamma\boldsymbol{v},\ \frac{i}{c}(\mathscr{E} - U)\right]$, since $m\gamma c = \frac{1}{c}m\gamma c^2 = \frac{\mathscr{E} - U}{c}$, as it is seen from Eqs. (5.21) and (5.57); using the relation $\vec{P}^2 = -m^2c^2$. Eq. (5.51) yields the Hamiltonian function for a particle in a conservative field:

$$\mathscr{E} = \mathscr{H} = c\sqrt{p^2 + m^2c^2} + U. \tag{5.59}$$

Just as in the case of a free particle, a 3-relativistic momentum can be expressed via the energy, velocity and potential energy:

$$\boldsymbol{p} = m\gamma\boldsymbol{v} = \frac{1}{c^2}mc^2\gamma\boldsymbol{v} = \frac{\mathscr{E} - U}{c^2}\boldsymbol{v}.$$

The transformation of the fourth component of the energy-momentum vector for a particle in a potential field shows that according to Eq. (5.42) (and making use of Eq. (5.10) as well) the total energy is equal to $\mathscr{E}' = mc^2\gamma' + U'$ in the frame K', as one would expect.

§ 5.6. The rest mass of a system. The binding energy. So far we have been considering mechanics of a "particle", i.e. the behaviour of a certain single unit. However, the "elementariness" (indivisibility) of a particle was, in fact, nowhere assumed, so that all conclusions may be transferred to more complex systems consisting of "subsystems".

The rest mass M of a complex system is defined in accordance with the general formula of Eq. (5.50):

$$M^2c^2 = \frac{E^2}{c^2} - \boldsymbol{P}^2, \tag{5.60}$$

where E is now the total energy of a system and $\boldsymbol{P}$ its total momentum.

Let us confine ourselves to the simplest systems consisting of individual particles. First suppose that the particles do not interact with one another. Then the energy of such a system will be the sum of energies of particles comprising the system:

$$E = \sum \mathscr{E}_i. \tag{5.61}$$

The additivity of energies signifies the absence of interaction. The total momentum of a system always adds up vectorwise from the momenta of individual particles, i.e. it is always additive:

$$\boldsymbol{P} = \sum_i \boldsymbol{p}_i. \tag{5.62}$$

In this case the rest mass of the system can be written as follows:

$$M^2c^2 = \frac{\left(\sum\limits_i \mathscr{E}_i\right)^2}{c^2} - \left(\sum_i \boldsymbol{p}_i\right)^2. \tag{5.63}$$

In order to find the relationship between the rest mass of a system and the rest masses of particles comprising the system, one should pass over to the reference frame in which the total momentum of the system is equal to zero: $\boldsymbol{P} = 0$. Then from Eq. (5.63) we obtain

$$M = \frac{\sum\limits_i \mathscr{E}_i}{c^2}. \tag{5.64}$$

We see that the rest mass of a system is expressed as the sum of energies of constituent particles (divided by c^2). But the energy of an individual particle can be represented according to Eq. (5.44) as the sum of its rest energy and kinetic energy

$$\mathscr{E}_i = m_i c^2 + T_i. \tag{5.65}$$

Then in accordance with Eq. (5.63) we get

$$M = \sum_i m_i + \frac{1}{c^2} \sum_i T_i. \tag{5.66}$$

Eq. (5.66) yields the important result: the rest mass of a system exceeds the sum of rest masses of constituent particles by the total kinetic energy of the particles (divided by c^2) estimated in the reference frame in which the total momentum of the system is equal to zero.

This way we reach a conclusion that in relativistic mechanics the rest mass of the system composed of non-interacting particles is not an additive quantity. Such a property of mass is uncommon in classical mechanics. It is tempting to bring in some new definition of mass for constituent particles, so that the rest mass of a system will add up from these new masses called "relativistic" sometimes. It is easy to see how this can be done. In the reference frame where $\boldsymbol{P} = 0$ we obtain, according to Eq. (5.64):

$$Mc^2 = \sum_i \mathscr{E}_i = \sum_i m_i \gamma_i c^2. \tag{5.67}$$

Consequently, one can write down

$$M = \sum m_i \gamma_i = \sum m_i^{rel}, \tag{5.68}$$

where $m_i^{rel} = m_i \gamma_i$ is a relativistic mass. That is how we realize additivity (which is by no means obligatory), but at the same time we clear the way for various delusions. In fact, the introduction of a relativistic mass for a particle creates an illusion that the increase of the particle's energy, or "relativistic mass", accompanying the increase of its velocity (momentum) is associated with the changes of the internal structure of the particle. But surely, there is no such thing at all (to make sure, one may just pass over to another frame without approaching the particle) As a matter of fact, the energy grows with the increase of velocity due to the special properties of the 4-space-time coming through in the Lorentz transformation.

In terms of the four-dimensional approach the term "mass" refers to the invariant norm of the 4-vector of energy-momentum. By bringing in a relativistic mass we actually apply the term "mass" (with an accuracy of a factor) to the time component of

the 4-vector of energy-momentum, which is, as we know, the energy. However, the energy and the rest mass that we intend to use represent essentially different physical concepts.

Energy is a relative quantity; it depends on the IFR in which the particle or the system of particles is considered. The rest mass remains the same in all IFRs; it is an absolute value of a 4-vector. The time component of a 4-vector (energy) coincides with its absolute value (rest mass) only when the spatial components of this 4-vector are equal to zero (which means that either the particle's momentum or the total momentum of a system of particles equals zero). And only in the case of the energy coinciding with the rest energy it is proportional to the rest mass (with the constant coefficient c^2).

Thus, we can ascribe the precise four-dimensional meaning to the momentum, energy and rest mass of a particle or a system, provided the first two quantities are treated as components of a 4-vector of energy-momentum, and the last quantity as a norm of the same vector. The methodical aspects of the problem are discussed in Supplement IV.

Let us examine now a system composed of interacting particles. Eq. (5.63) remains valid, of course. Eq. (5.61) should be, however, replaced by

$$E=\sum_i \mathscr{E}_i+U, \tag{5.69}$$

where U denotes the interaction energy of particles. This energy is defined as the work required to disjoint the system into "initial" non-interacting parts. In a stable system $U<0$ since a "stable equilibrium" state is characterized by a minimum of energy. In such a system the quantity U is referred to as a *binding energy*. Although the explicit analytical expression for the interaction energy is often rather difficult to write out (see § 5.8), its magnitude can be estimated. From Eq. (5.60) we obtain for the reference frame in which $\boldsymbol{P}=0$:

$$M=\frac{\sum_i \mathscr{E}_i+U}{c^2}, \tag{5.70}$$

or expressed otherwise

$$M=\sum_i m_i+\frac{1}{c^2}\sum_i T_i+\frac{U}{c^2}, \tag{5.71}$$

where we made use of the relation $\mathscr{E}_i=m_ic^2+T_i$ which is correct for every individual particle.

If the condition $\sum_i T_i \ll U$ is met, i.e. the total relativistic kinetic energy of particles is small, then

$$M = \sum_i m_i + (U/c^2). \tag{5.72}$$

It is evident from Eq. (5.72) that in the system of interacting particles the difference

$$\Delta M = \sum_i m_i - M, \tag{5.73}$$

customarily called a *mass defect*, is always different from zero. In a stable system $U < 0$ and $\Delta M > 0$. From the mass defect one can calculate the binding energy:

$$U = \Delta M \cdot c^2. \tag{5.74}$$

Such a calculation is meaningful only when binding energies are substantial. Precisely such a case is realized in atomic nuclei. It is well known that atomic nuclei are very stable, this being the evidence of their considerable binding energy. Atomic nuclei are composed of protons and neutrons, with each nucleus possessing a quite definite number of protons and neutrons. The masses of protons and neutrons in a free state (outside a nucleus) can be experimentally determined. The mass of any atomic nucleus can also be experimentally found. The difference between the summarized mass of free protons and neutrons comprising the nucleus and the measured mass of the nucleus yields the mass defect and, according to Eq. (5.74), the binding energy. Exactly in this manner the binding energies of nuclei are found in atomic physics.

Let us write out separately expressions for ultra-relativistic particles ($v \approx c$). In this case Eq. (5.51′) holds:

$$\mathscr{E} = cp \tag{5.75}$$

and consequently from Eq. (5.51) we shall obtain $m = 0$. However, for two or more particles we shall obtain from Eq. (5.63)

$$M^2c^2 = \left(\sum_i p_i\right)^2 - \left(\sum_i \mathbf{p}_i\right)^2 \neq 0, \tag{5.76}$$

since $\sum_i \mathscr{E}_i = c \sum_i p_i$. This points out that the rest mass of a system composed of particles with the zero rest mass is not equal to zero. There is nothing surprising in this since rest masses do not add up.

Finally, a few words about "composite" subsystems. When determining a rest mass of a complex system according to Eqs.

(5.70) and (5.64), one should take a total energy of the system. Assume that the system involves an electromagnetic field as well. Designating the energy of the electromagnetic field by W, we obtain from Eq. (5.70):

$$M = \sum_i m_i + \frac{1}{c^2} \sum_i T_i + \frac{U}{c^2} + \frac{W}{c^2}. \qquad (5.77)$$

From this it is inferred that the energy of an electromagnetic field, as any other energy, makes its contribution to the rest mass of a system.

§ 5.7. Some problems of relativistic mechanics of a particle. Within the framework of a given inertial frame of reference there is no need to resort to four-dimensional relations; it is sufficient to use the three-dimensional Eq. (5.26) and also Eq. (5.31). We shall recall that the general appearance of the second law of Newton remains invariable and only a momentum and energy of a particle are defined otherwise. This new definition, however, changes substantially the properties of the solution of the problem as compared to that obtained for the same problem from classical mechanics equations. In particular, the solution of any problem of relativistic mechanics does not permit of obtaining the velocity of a particle exceeding that of light. Some other distinguishing features come up as well; to elucidate them, we shall consider a few problems, solving them with the aid of both the classical and relativistic motion equations.

Since the equations obtained will not be needed afterwards, an individual numeration for each problem is adopted in this section.

1. The elementary solution of the problem of the unidimensional motion under the action of a constant force. The motion equation takes the form (a classical equation on the left side and a relativistic one on the right side):

$$\frac{d}{dt}(mv) = F; \quad \text{(a)} \quad \Big| \quad \frac{d}{dt}(m\gamma v) = F, \quad \text{(b)} \qquad (1)$$

where $F = \text{const}$. Integrating with the initial condition $v = 0$ at $t = 0$, we obtain

$$mv = Ft; \quad \text{(a)} \quad \Big| \quad m\gamma v = Ft \quad \left(\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}\right). \quad \text{(b)} \qquad (2)$$

A velocity as a function of time is found from Eqs. (2a, b) algebraically:

$$v_{cl} = \frac{Ft}{m}; \quad \text{(a)} \quad \Big| \quad v_{rel} = \frac{Ft/m}{\sqrt{1 + \left(\frac{Ft}{mc}\right)^2}} = \frac{v_{cl}}{\sqrt{1 + \left(\frac{v_{cl}}{c}\right)^2}}. \quad \text{(b)} \qquad (3)$$

We shall discuss these results in Problem II as it will become evident that Problem II is just a variant formulation of Problem I.

II. The rectilinear and uniformly accelerated motion of a particle. If a particle moves along the x' axis ($v'_y = v'_z = 0$) in the frame K', then $v_y = v_z = 0$ in any other IFR according to Eq. (3.26). Let us consider a motion of a particle along the common x, x' axis with a constant acceleration. If the acceleration of the particle does not vary, the acceleration components in the reference frame in which this particle is at rest (the proper reference frame) are (w_0, 0, 0, 0). The quantity w_0 is a customary three-dimensional acceleration directed along the x' axis. The square of a 4-vector acceleration is an invariant, so that the following condition must hold in all reference frames for a uniformly accelerated motion *:

$$\left(\frac{du_i}{d\tau}\right)^2 = w_0^2, \tag{1}$$

where w_0 is the magnitude of the three-dimensional acceleration in the proper reference frame. Surely, this condition differs from the requirement $\dot{v} = 0$ **. The 4-velocity components for a unidimensional motion in any reference frame acquire the form $\vec{V}(u_1 = \gamma v_x,\ 0,\ 0,\ ic\gamma)$ whence, according to Eq. (5.23), it follows that $\mathfrak{F}_2 = \mathfrak{F}_3 = 0$. Consequently, the two equations that remain in an arbitrary IFR are

$$m\frac{du_1}{d\tau} = \gamma F, \quad m\frac{du_4}{d\tau} = \frac{i}{c}\gamma F v.$$

We denoted here $F_x = F$ and $v_x = v$. Eq. (1) will take the form

$$\left(\frac{du_i}{d\tau}\right)^2 = \frac{1}{m^2}\left(\gamma^2 F^2 - \frac{1}{c^2}\gamma^2 F^2 v^2\right) = \frac{F^2}{m^2} = w_0^2 = \text{const.}$$

The motion equation will be written as

$$\frac{du_1}{d\tau} = \gamma w_0, \quad \text{or} \quad \frac{du_1}{dt} = w_0. \tag{2}$$

In the three-dimensional notation

$$\frac{d}{dt}\left(\frac{v}{\sqrt{1-\frac{v^2}{c^2}}}\right) = w_0, \quad \text{or} \quad \frac{v}{\sqrt{1-\frac{v^2}{c^2}}} = w_0 t + C. \tag{3}$$

---

* In this chapter we make use of the 4-space-time whose fourth coordinate is imaginary.

** If $w_0 = \text{const}$, $\dot{v}$ is variable. In classical mechanics an object acquires a constant acceleration under the action of a constant force, whereas in relativistic dynamics a constant force imparts a constant acceleration to an object only in an instantaneous co-moving reference frame.

Solving Problem I, we got convinced, however, that these equations could be obtained at once if we substituted the constant force $F = mw_0$ into the relativistic motion equation. But w_0 is not the same as $\dot{v}$; this can be inferred from Eq. (2).

If the initial condition is such that the velocity v is equal to zero at $t = 0$, then $C = 0$ and expressing the velocity through w_0, we obtain from Eq. (3)

$$v = \frac{w_0 t}{\sqrt{1 + \frac{w_0^2 t^2}{c^2}}} = \frac{v_{cl}}{\sqrt{1 + \left(\frac{v_{cl}}{c}\right)^2}}, \quad v = \frac{dx}{dt}, \tag{4}$$

where v_{cl} denotes $w_0 t$. Integrating the latter relation at the initial conditions, $x = 0$, $t = 0$, we obtain the following expression:

$$x = \frac{c^2}{w_0}\left(\sqrt{1 + \left(\frac{v_{cl}}{c}\right)^2} - 1\right). \tag{5}$$

The solution of the classical motion equation for a constant force and the identical initial conditions takes the form

$$v_{cl} = w_0 t, \quad x = w_0 t^2/2.$$

If the classical velocity grows indefinitely with time, it follows from Eq. (4), due to the obvious inequality

$$\frac{x}{\sqrt{A + \frac{x^2}{a^2}}} = \frac{x}{\sqrt{Aa^2 + x^2}}\, a < a, \quad \text{if} \quad A > 0,$$

that the relativistic velocity always remains less than c as it, in fact, must be in accordance with the principle of ultimate velocity of signal transmission. The relativistic equations (4) and (5) for the velocity v and coordinate x turn into classical ones at $v_{cl}/c \ll 1$. If one rewrites Eq. (4) as $v = c/\sqrt{1 + c^2/w_0^2 t^2}$, it becomes evident that $v \to c$ when $t \to \infty$.

Let us find the relationship between the coordinate time t and the proper time τ of a particle. If one chooses the common origin of the time count $t_0 = \tau_0 = 0$, then

$$\tau = \int_0^t \sqrt{1 - \frac{v^2}{c^2}}\, dt = \int_0^t \sqrt{1 - \frac{w_0^2 t^2/c^2}{1 + \frac{w_0^2 t^2}{c^2}}}\, dt =$$

$$= \frac{c}{w_0} \operatorname{Arcsinh} \frac{w_0 t}{c} = \frac{c}{w_0} \ln\left(\frac{w_0 t}{c} + \sqrt{1 + \frac{w_0^2 t^2}{c^2}}\right) \tag{6}$$

(see Eq. (3.16)).

Ignoring unity in the radicand as compared to w_0t/c, we obtain, when $t \to \infty$,

$$\tau \sim \frac{c}{w_0} \ln \frac{2w_0t}{c}.$$

We see that the proper time of an object moving with a uniform acceleration flows substantially slower than the time in a "motionless" reference frame relative to which the motion is considered.

There remains, of course, the physical question as to what clock counts the particle's proper time given by Eq. (6) since the relation $d\tau = (1/\gamma)\,dt$ pertains to a clock moving uniformly and rectilinearly. We examined this problem in detail in § 3.3.

In conclusion we should note that a relativistic uniformly accelerated rectilinear motion is also called hyperbolic since the time dependence of the path travelled (see Eq. (7) below) represents hyperbola in terms of geometry. If a charged particle moves in a uniform and constant electric field, or a heavy particle in a uniform and constant gravitational field, the motion is hyperbolic. Let us write out finally the principal equations describing a hyperbolic motion:

$$x(t) = \frac{c^2}{w_0}\left[\sqrt{1 + \left(\frac{w_0t}{c}\right)^2} - 1\right],$$
$$v = \frac{w_0t}{\sqrt{1 + (w_0t/c)^2}}, \quad \dot{v} = \frac{dv}{dt} = \frac{w_0}{\sqrt{1 + (w_0t/c)^2}}. \tag{7}$$

From the expression for the time derivative of the velocity one may see the difference between a relativistic and non-relativistic "constant" acceleration.

III. The motion of a charged particle in a constant uniform electric field. Let us choose the following initial conditions: at the moment $t = 0$ the coordinates of a charged particle $x_0 = y_0 = 0$, and its velocity v_0 is perpendicular to the field $\boldsymbol{E}$. This corresponds to the problem about a particle flying into a charged capacitor parallel to its plates (Fig. 5.1). Let us direct the x axis along $\boldsymbol{E}$ and the y axis along v_0. Then the motion of the particle will take place in the plane (x, y). As far as it is possible, we shall not discriminate between classical and relativistic equations of motion, writing them in the common form:

$$\frac{dp}{dt} = e\boldsymbol{E},$$

where $\boldsymbol{F} = e\boldsymbol{E}$ is the force exerted by the electric field on the charged particle. In components:

$$\dot{p}_x = eE, \quad \dot{p}_y = 0.$$

Whence the momentum is found by integration:

$$p_x = eEt + p_{0x}, \quad p_y = p_{0y}.$$

But according to the initial conditions both in classical and rela-

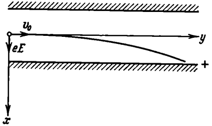


Fig. 5.1. An electron flying into the uniform electric field of a capacitor is at the origin of coordinates at the moment $t = 0$ The force exerted by the field is directed along the x axis. The initial velocity of an electron v_0 is directed along the y axis The classical solution of the problem coincides with that of a problem of the motion of a heavy point thrown horizontally in the gravitational field at the velocity v_0

tivistic cases $p_x = 0$ and $p_{0y} = p_0$ at $t = 0$. Consequently, one may write

$$p_x = eEt, \quad p_{0y} = p_0.$$

From here on we should take into account the difference in definitions of a momentum:

classical:

$$p = mv.$$

We have

$$v_x = \frac{dx}{dt} = \frac{eEt}{m}, \qquad (1)$$

$$v_y = \frac{dy}{dt} = v_0.$$

Integrating, we get

$$x = \frac{eE}{m}\frac{t^2}{2} + x_0,$$

$$y = v_0 t + y_0.$$

But according to the initial conditions $x_0 = y_0 = 0$ at $t = 0$, we finally obtain

$$x = \frac{eE}{m}\frac{t^2}{2}, \quad y = v_0 t. \qquad (*)$$

relativistic:

$$p = m_\text{v} v.$$

In this case

$$p^2 = p_x^2 + p_y^2 = (eEt)^2 + p_0^2.$$

The particle's energy $\mathscr{E}$ will be defined as

$$\mathscr{E} = c\sqrt{p^2 + m^2c^2} =$$

$$= \sqrt{m^2c^4 + p_0^2c^2 + (ceEt)^2} =$$

$$= \sqrt{\mathscr{E}_0^2 + (ceEt)^2}.$$

Using Eq. (5.55)

$$v_{rel} = \frac{c^2}{\mathscr{E}}p, \qquad (1')$$

The velocity of a particle

$$v^2 = v_x^2 + v_y^2 = \left(\frac{eE}{m}\right)^2 t^2 + v_0^2$$

increases indefinitely with time. The particle's path obtained from Eq. (*) by elimination of the time t is represented by a parabola

$$x = \frac{eE}{2mv_0^2} y^2. \qquad (2)$$

we obtain

$$v_x^{rel} = \frac{dx}{dt} = \frac{p_x c^2}{\mathscr{E}} = \frac{c^2 eEt}{\sqrt{\mathscr{E}_0^2 + (ceEt)^2}} = \frac{v_x}{\sqrt{\left(\frac{\mathscr{E}_0}{mc^2}\right)^2 + \left(\frac{v_x}{c}\right)^2}}, \qquad (2')$$

where v_x^{rel} is defiined according to Eq. (1'). It follows from Eq. (2') that, as in the previous section, v^{rel} is always less than c because

$$v_y^{rel} = \frac{dy}{dt} = \frac{p_0 c^2}{\sqrt{\mathscr{E}_0^2 + (ceEt)^2}} \qquad (3')$$

diminishes with time. Integrating Eq. (2'), we obtain with regard to the initial conditions

$$x = \frac{1}{eE}\sqrt{\mathscr{E}_0^2 + (ceEt)^2} - \frac{\mathscr{E}_0}{eE}$$

Integrating Eq. (3') and taking into account the initial conditions, we get

$$y = \frac{p_0 c}{eE} \operatorname{Arcsinh} \frac{ceEt}{\mathscr{E}_0}.$$

Eliminating t from the expressions for x and y, we have

$$x = \frac{\mathscr{E}_0}{eE}\left(\cosh \frac{eEy}{cp_0} - 1\right).$$

Thus, when a classical path was a parabola, a relativistic one turned out to be a catenary curve. In the case of $v \gg c$ a catenary curve, however, turns into a parabola. Indeed, if $v/c \ll 1$, we have $\gamma \approx 1$, $p_0 = mv_0$, and $\mathscr{E}_0 = mc^2$. Besides, at small x one may assume $\cosh \theta \approx 1 + \theta^2/2!$, whence

$$x = \frac{\mathscr{E}_0}{eE} \frac{e^2E^2y^2}{2p_0^2c^2} = \frac{mc^2}{eE} \frac{(eE)^2 y^2}{2m^2v_0^2c^2} = \frac{eE}{2mv_0^2} y^2,$$

which is the parabola of Eq. (2). This example, as well as all subsequent ones, shows that problems of relativistic mechanics

do not require any velocity dependence of mass: the solution is obtained by integrating a motion equation.

IV. The motion of a charged particle in a constant uniform magnetic field. The classical and relativistic motion equations for a charged particle in a magnetic field

$$\frac{d\boldsymbol{p}}{dt}=e[\boldsymbol{v}\boldsymbol{B}] \qquad (1)$$

are the same not only in their appearance. The point is that a magnetic field does not perform any work on a charged particle and the energy of the particle remains constant (see Eqs. (5.20) and (5.31)); of course, the expressions for energy in classical and relativistic cases are different. Using the relativistic relation

$$\boldsymbol{p}=\frac{\mathscr{E}}{c^2}\boldsymbol{v},$$

where $\mathscr{E}$ = const, Eq. (1) can be rewritten in the form

$$\frac{d\boldsymbol{v}}{dt}=\frac{ec^2}{\mathscr{E}}[\boldsymbol{v}\boldsymbol{B}], \qquad (2)$$

whereas from the classical definition of a momentum $\boldsymbol{p}=m\boldsymbol{v}$ it follows that

$$\frac{d\boldsymbol{v}}{dt}=\frac{e}{m}[\boldsymbol{v}\boldsymbol{B}]. \qquad (3)$$

Therefore, relativistic and classical motion equations (2) and (3) are distinguished only by the constants standing in front of the vector product. We shall recall how Eq. (2) or (3) is solved. Orient the z axis along a magnetic field. Then $\boldsymbol{B}=B\boldsymbol{k}$. Denote the constant factors appearing in front of the product $[\boldsymbol{v}\boldsymbol{k}]$ in Eqs. (2) and (3) as follows:

$$\omega_{cl}=\frac{eB}{m},\quad \omega_{rel}=\frac{ec^2B}{\mathscr{E}}=\frac{eB}{m\gamma}=\frac{1}{\gamma}\omega_{cl}=\omega_{cl}\sqrt{1-\beta^2}. \qquad (4)$$

Now for definiteness let us solve Eq. (2). To define the vector product $[\boldsymbol{v}\boldsymbol{k}]$, rewrite Eq. (2) in components:

$$\dot{v}_x=\omega v_y,\quad \dot{v}_y=-\omega v_x,\quad \dot{v}_z=0. \qquad (5)$$

It is expedient to resort to a complex variable in the plane (v_x, v_y). Multiplying the second relation of (5) by the number i and adding the result to the first one, we obtain $\frac{d}{dt}(v_x+iv_y)=$ $=-i\omega(v_x+iv_y)$.

This equation can readily be integrated:

$$v_x+iv_y=ae^{-i\omega t},$$

where a is a complex constant. If it is written in the form $a = v_{0t}e^{-i\alpha}$ with the real v_{0t} and α, the solution takes the following form:

$$v_x + iv_y = v_{0t}e^{-i(\omega t+\alpha)}. \tag{6}$$

Obviously, v_{0t} is a modulus of the complex number in the left-hand side of Eq. (6);

$$v_{0t}^2 = v_x^2 + v_y^2.$$

Consequently, the magnitude of the particle's velocity remains constant in the plane (x, y). Eq. (6) can be rewritten in the form

$$\frac{d}{dt}(x + iy) = v_{0t}e^{-i(\omega t+\alpha)},$$

permitting of direct integration:

$$x + iy = \frac{v_{0t}}{\omega}e^{-i(\omega t+\alpha-\pi/2)}. \tag{7}$$

Recalling the geometric representation of the complex quantity $w = x + iy = re^{i\varphi}$, we see that a particle remains permanently on a circumference of a constant radius $r = v_{0t}/\omega$, while the angle between its radius vector and the x axis increases evenly with time: $\varphi = \omega t + \text{const}$. This means that the projection of the particle's motion on the plane (x, y) is a uniform motion along a circumference of the radius

$$r = \frac{v_{0t}}{\omega} = \frac{v_{0t}\mathcal{E}}{ec^2B} = \frac{p_t}{eB}, \tag{8}$$

where p_t is the projection of a momentum on the plane (x, y) and ω is an angular velocity. As to the motion along the z axis, it follows from the third relation (5) that

$$z = z_0 + v_{0z}t. \tag{9}$$

From Eqs. (8) and (9) it follows that a charged particle in a uniform magnetic field moves along a helical line whose axis coincides with the direction of a magnetic field, and whose radius is determined according to Eq. (8). The velocity of a particle is constant, as it should be in a magnetic field. If at the initial moment the velocity of a particle in the direction of a magnetic field is equal to zero ($v_{0z} = 0$), the particle moves along the circumference in the plane perpendicular to the field.

The quantity ω_{rel} defines the cyclic frequency of rotation of the particle's projection on the plane (x, y) perpendicular to the direction of a magnetic field. This frequency is referred to as *cyclotronic*. As we have seen, $\omega_{cl} = \gamma\omega_{rel}$, i.e. the cyclotronic frequency of relativistic particles is less than that of non-relativistic ones. At small velocities $\gamma \to 1$ and $\omega_{cl} \to \omega_{rel}$.

In conclusion let us consider an acceleration gained by a particle in an electromagnetic field in terms of classical and relativistic mechanics. From the general motion equation

$$\frac{d\boldsymbol{p}}{dt}=e\{\boldsymbol{E}+[\boldsymbol{vB}]\}$$

we obtain in the classical case ($\boldsymbol{p}=m\boldsymbol{v}$)

$$\dot{\boldsymbol{v}}_{cl}=\frac{e}{m}\{\boldsymbol{E}+[\boldsymbol{vB}]\}.$$

In order to obtain an acceleration in the relativistic case we shall make use of Eq. (5.55), whence

$$\frac{d\boldsymbol{p}}{dt}=\frac{\boldsymbol{v}}{c^2}\frac{d\mathscr{E}}{dt}+\frac{\mathscr{E}}{c^2}\frac{d\boldsymbol{v}}{dt}.$$

According to Eq. (5.31) $d\mathscr{E}/dt=\boldsymbol{Fv}=e\boldsymbol{Ev}$, and according to Eq. (5.32) $\mathscr{E}/c^2=m\gamma$, so that (cf. Eq. (5.38))

$$\dot{\boldsymbol{v}}_{rel}=\frac{e}{m\gamma}\left\{\boldsymbol{E}+[\boldsymbol{vB}]-\frac{\boldsymbol{v}}{c^2}(\boldsymbol{Ev})\right\}=$$

$$=\frac{1}{\gamma}\dot{\boldsymbol{v}}_{cl}-\frac{e}{m\gamma c^2}\boldsymbol{v}(\boldsymbol{Ev})=\frac{1}{\gamma}\dot{\boldsymbol{v}}_{cl}-\frac{e}{\mathscr{E}}\boldsymbol{v}(\boldsymbol{Ev}).$$

The second term in the last link of the equation can be regarded as an emergence of a certain friction (proportional to the velocity); owing to this, one may realize in qualitative terms that the acceleration of a particle decreases sharply as the particle's velocity approaches that of light. It is obvious, of course, that $\dot{\boldsymbol{v}}_{cl}=$ $=\dot{\boldsymbol{v}}_{rel}$ with an accuracy to within β^2. The motion of a charged particle in constant electric and magnetic fields is presented in detail in [9], § 22; we should only point out here that in the case of crossed (mutually perpendicular) fields for which $E^2-c^2B^2\neq$ $\neq 0$ holds (cf. § 6.5), a transition to a certain inertial frame of reference may eliminate one of the fields and leave either an electric or a magnetic field. Then in this reference frame one may utilize the results obtained here.

V. The reaction motion in relativistic mechanics. As in the previous problems, we shall be examining classical and relativistic cases simultaneously. As an example, let us consider the motion of a rocket which (together with the ejected gas) can be treated as a closed system. We shall recall that the rocket propulsion is brought about due to the fact that during each time interval it ejects a certain amount of substance at a definite velocity with respect to the rocket. In accordance with the momentum conservation law the rocket shell with the left-over fuel acquires a momentum in the direction opposite to the direction in which the gas jet is ejected. Both in classical and in relativistic cases the

problem is easier to solve in the inertial frame of reference co-moving with the rocket. Since the velocity of the rocket changes, we deal with the instantaneous co-moving frame.

Let at the moment t a rocket mass (fuel and container) be $M(t)$ and a velocity $V(t)$; during the time dt the rocket engine ejects a mass dM of gas at a constant velocity *relative to the rocket*, v. Write out the momentum conservation law in a co-moving frame (a unidimensional case):

$$(M - dM)\,dV - v\,dM = 0, \quad (1)$$

with $dM > 0$. Ignoring infinitesimal values of the second order $(dM \cdot dV)$, we get

$$\frac{dM}{M} = -\frac{dV}{v}. \quad (2)$$

Eq. (2) is easily integrated ($v = \text{const}$):

$$V = v \ln \frac{C}{M},$$

where C is an integration constant. Choosing the initial conditions: at $t=0$ $V=0$, $M(0) = M_0$, we finally obtain

$$V = v \ln \frac{M_0}{M}. \quad (3)$$

Eq. (3) determines the velocity $V(t)$ of the rocket as a function of the velocity of ejected gases and the change of the rocket mass (the mass of the burnt fuel is equal to $M_0 - M$).

There is, however, one noteworthy feature in this derivation. At any moment of time the rocket has different co-moving inertial frames. The velocity increment in Eq. 2 during the time interval dt is

In a relativistic case one should consider not a velocity increment, but a velocity parameter increment, since velocity parameters add up (are additive), while velocities do not (see Eq. (3.37)).

Write out the relation between the velocity and the velocity parameter θ (see Eqs. (2.27), (2.28)):

$$\tanh\theta = \beta, \quad \cosh\theta = \gamma, \quad \sinh\theta = \gamma\beta, \quad (5)$$

when examining the velocity of ejected gases, and

$$\tanh\theta = \mathrm{B}, \quad \cosh\theta = \Gamma, \quad \sinh\theta = \Gamma \mathrm{B}, \quad (6)$$

when dealing with the velocity of the rocket. Obviously,

$$\mathscr{E} = mc^2 \cosh\theta, \quad p = m \sinh\theta. \quad (7)$$

Suppose that during the time interval dt the mass dM is ejected at the velocity $\beta = v/c$. The velocity of the rocket after the ejection of the mass dM increases by $d\mathrm{B} = dV/c$. We shall express the increment $d\mathrm{B}$ via the velocity parameter increment (see Eq. (6)):

$$d\mathrm{B} = \tanh(d\theta), \quad (8)$$

Now write down the laws of momentum and energy conservation. In a co-moving frame the momentum of a rocket "before the ejection" is equal to zero ($\mathrm{B} = 0$). After the ejection of the mass dM the rocket acquires the momentum $(M +$

written for different inertial frames of reference. In transition from one inertial frame of reference to another all velocities (and velocity increments) add up in classical mechanics. Therefore, it is immaterial that Eq. (2) refers to different IFRs; the final velocity can be obtained by summing (integrating) velocity increments over the total time interval during which the velocity of the rocket varies from 0 to V:

$$\int_{M_0}^{M} \frac{dM}{M} = -\frac{1}{v}\int_{0}^{V} dV. \quad (4)$$

This is just the method that was used in derivation of Eq. (3). Hence it is clear how important in this derivation is the additivity of velocities in transition from one IFR to another.

In the foregoing calculations the mechanical energy conservation law was not used because, on the one hand, it is insufficient (the thermal energy is also significant here), and on the other, it is not necessary when the velocity of a rocket is calculated.

$+ dM$) sinh ($d\theta$), while the momentum of the mass dM directed oppositely is equal to dM sinh(θ). Hence,

$$-dM \sinh(\theta) + \\ + (M + dM)\sinh(d\theta) = 0. \quad (9)$$

Relativistic mechanics makes it possible to take into account any energy transformations, so that here we can write the energy conservation law as well:

$$dMc^2\cosh(\theta) + \\ + (M + dM)c^2\cosh(d\theta) = Mc^2. \quad (10)$$

But we may regard $d\theta$ as a small quantity, so that $\sinh(d\theta) \simeq d\theta$, $\cosh(d\theta) \simeq 1$; consequently, ignoring infinitesimal values of the second order $dM \cdot d\theta$, we obtain from Eq. (9)

$$dM \sinh(\theta) = M\, d\theta, \quad (11)$$

and from Eq. (10)

$$\cosh\theta = -1. \quad (12)$$

Dividing termwise Eq. (11) by Eq. (12), we get

$$d\theta = -\frac{dM}{M}\tanh\theta,$$

or

$$d\theta = -\beta\frac{dM}{M}, \quad (13)$$

where $M = M(t)$ is the mass of the rocket and fuel at the moment t. Since in relativistic mechanics a velocity parameter is an additive value, the

final magnitude of a velocity parameter can be found by integration:

$$\theta = \beta \ln \frac{M_0}{M}. \qquad (14)$$

where M_0 is the mass of the rocket at the moment when its velocity is equal to zero.

Eq. (14) also determines implicitly the velocity of the rocket at the moment when the mass of the burnt fuel is equal to $M_0 - M$.

It is easy to see that if a rocket moves at a non-relativistic velocity, Eq. (14) turns into Eq. (3). In fact, in this case $B \ll 1$ and, consequently, $\tanh\theta = B \ll 1$, whence $\tanh\theta \approx \theta$ and Eq. (14) coincides with Eq. (3). Just as in all solutions of relativistic mechanics, the velocity of a rocket cannot exceed that of light c. Even if we manage to burn all the mass of the rocket together with the fuel, i.e. $M \to M_0$, $\ln(M_0/M) \to \infty$. From Eq. (14) it only follows that $\theta \to \infty$ (with the maximum value of β being, of course, equal to 1). But $B = \tanh\theta$ and at $\theta \to \infty$ $\tanh\theta \to 1$, i.e. $V \to c$.

Surely, the higher the ejection velocity, the more effective is the rocket performance. Can the ejection velocity be made equal to c, i.e. $\beta = 1$? It can be, provided light serves as a reaction gas: only photons and neutrinos can move at the velocity c. These two sorts of particles are peculiar in that their rest masses are equal to zero (§ 7.6). The zero mass, however, is seen directly from Eq. (10) in this case. As $v \to c$ a velocity parameter satisfies the condition $\tanh\theta = \beta \to 1$. But with $\tanh\theta \to 1$ $\cosh\theta \to \infty$, and in order to satisfy Eq. (10) at a finite value of M, it is necessary that $dM = 0$.

The more detailed analysis of photon rocket capabilities shows its inapplicability for long-range space flights (see [11]).

VI. Colliding beams. The progress in nuclear physics depends essentially on how high are the energies of interaction between elementary particles that could be observed. There have been two sources of high-energy particles so far: cosmic rays and accelerators. The accelerator designers are still very far from mastering the energies which particles in cosmic rays possess, while the systematic research in high-energy physics has been restricted by the range of energies covered by accelerators. Accelerators are complex and expensive machines whose construction continues

for years and whose cost amounts to an appreciable portion of a national budget of any highly developed country.

Suppose that in a laboratory reference frame particles are accelerated to the energy $\mathscr{E}$. We are to succeed in colliding these particles with other particles of the same type (for example, we examine proton-proton collisions). The proton beam, possessing the total energy $\mathscr{E}$ in a laboratory reference frame, can be directed onto a target containing hydrogen in which protons are practically motionless. It is sufficient to consider a collision of one bombarding and one resting proton. Then the energy of the system consisting of these two particles is equal to $\mathscr{E} + m$ (we assume $c = 1$ in this section). The following question arises: is it possible to increase essentially the interaction energy by making two beams, each consisting of particles possessing the energy $\mathscr{E}$ (in a laboratory frame), move toward each other? How much higher will be the "useful" interaction energy? For the sake of diversity we shall be treating this problem in rather unusual time units, which are light metres (see Chapter 2).

To simplify the problem, we shall not speak of beams any more, but consider only two particles. The maximum useful energy (spent on the generation of new particles, nuclear reactions, heating of a substance etc.) can be evaluated in the frame of the centre of inertia, for it is in this frame that the internal energy of the system is calculated (naturally, the motion of the system as a whole is "useless" from our point of view). Let us consider two particles *1* and *2* possessing equal energies (velocities) in a laboratory frame K, and moving toward each other. This frame will be for them a frame of the centre of inertia, and the total energy of particles in this frame provides the useful energy under discussion. This total energy is equal to $2T + 2m$, where T is the kinetic energy of each particle, and $2m$ the rest energy of particles (in the adopted time units $c = 1$). Let us find out how the same collision would look in the frame K' in which particle *1* is at rest. This is just the picture of a particle impinging against a target. We shall consider the energy of particle *2* calculated in the frame where particle *1* is at rest. Let us make the appropriate recalculation according to Eq. (5.43). (Do not forget that the same collision is being considered in another reference frame.) Designate the momentum and energy of particles *1* and *2* in the frame K by $(p, \mathscr{E})$ and $(-p, \mathscr{E})$. In the frame K' the momentum and energy of particle *1* are equal to $p_1' = 0$, $\mathscr{E}_1^0 = m$. Fig. 5.2 illustrates the frames K and K' and the velocities of particles in the frame K. The frame K' is the proper frame of particle *1*, and according to Eq. (5.49)

$$\mathscr{E}_1 = \Gamma \mathscr{E}_1^0 = \Gamma m = \gamma m \qquad (1)$$

(in our case $\Gamma = \gamma$ since K' is associated with particle *1*). It follows immediately from Eq. (1) that

$$\gamma = \frac{\mathscr{E}_1}{m} = \frac{T+m}{m} \approx \frac{T}{m}.$$

We shall consider relativistic velocities of particles when $\beta \approx 1$; consequently

$$p = \beta \mathscr{E} \approx \mathscr{E}.$$

In our case $\beta = \mathrm{B} \approx 1$, whence the quantity $\Gamma\mathrm{B}$ entering in the transformation of energy is approximately equal to Γ. Now it is

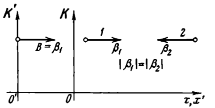


Fig. 5.2. Two points move in a laboratory frame at equal and oppositely directed velocities. Particle *1* is at rest in the frame K'.

easy to obtain the formula for transformation of energy of particle *2* on transition from the reference frame K to K':

$$\mathscr{E}'_2 = \Gamma(\mathscr{E}_2 - \mathrm{B}p_2) = \Gamma(\mathscr{E} + \mathrm{B}p) \approx \Gamma\mathscr{E} + \Gamma p \approx 2\Gamma\mathscr{E}.$$

$2\mathscr{E}$ is the energy realized in a head-on collision and accurate to a doubled rest energy of particles (which may be neglected at relativistic velocities). To realize this energy with particle *1* resting, one needs the energy Γ times higher. This calculation shows the advantage that we gain when we use collision beams.

The same result, however, can be obtained by simpler means. We shall demonstrate that the recalculation of energy according to Eq. (5.43) is equivalent to the calculation of energy according to the formula $\mathscr{E} = m\gamma$ into which the relativistic expression of the relative velocity of particles *1* and *2* is substituted.

Thus, let the momentum of particle *2* in the frame K be equal to $p_2 = p_x = -m\gamma\beta$ and energy $\mathscr{E}_2 = m\gamma$. We shall calculate the energy of particle *2* in the frame K' in which particle *1* is at rest.

According to Eq. (5.43)

$$\mathscr{E}'_2 = \Gamma(\mathscr{E} - \mathrm{B}p_2) = \Gamma(m\gamma + \mathrm{B}m\gamma\beta) = \Gamma m\gamma(1 + \beta\mathrm{B}),$$

since in our case $\gamma = \Gamma$ (in the frame K both particles have the same velocities). However, from Eq. (5.10) it follows (do not

forget that we have $\beta < 0$, see Fig. 5.2) that

$$\Gamma\gamma(1 + B\beta) = \gamma';$$

therefore

$$\mathscr{E}'_2 = m\gamma',$$

where γ' is determined for the velocity of particle *2* in the frame K' (i.e. the velocity of particle *2* relative to particle *1*). Calculate the relative velocity of particles *1* and *2*. We have

$$\beta'_2 = \beta'_x = \frac{\beta_x - B}{1 - \beta_x B}.$$

In the frame K particle *2* has the velocity $-\beta$ and particle *1* the velocity β, and consequently

$$\beta'_x = -\frac{2\beta}{1 + \beta^2}.$$

This is just the relative velocity of particle *2*. Find now γ':

$$\gamma' = \frac{1}{\sqrt{1 - \beta_2'^2}} = \frac{1}{\sqrt{1 - \left(\frac{2\beta}{1 + \beta^2}\right)^2}} = \frac{1 + \beta^2}{1 - \beta^2} = \gamma^2(1 + \beta^2).$$

Consequently,

$$\mathscr{E}'_2 = m\gamma' = m\gamma \cdot \gamma(1 + \beta^2) \approx 2\Gamma\mathscr{E},$$

for $\beta \approx 1$ and $\gamma = \Gamma$. We have obtained the same result again, as it should be.

§ 5.8. The conservation laws of relativistic mechanics. So far we have discussed the energy and momentum conservation laws for a material point, and now we have to dwell on the conservation laws for a system of n material points. The problem of conservation laws has two aspects. The first one involves the acquisition of relativistic laws of conservation in terms of a given IFR. The second aspect pertains to the examination of the behaviour of the quantities remaining constant on transition from one inertial frame to another. Both of these problems are solved by obvious methods for a system of non-interacting particles; in the case of interacting particles these problems are very complicated.

Let us begin with a system of n non-interacting particles. The motion equations and energy changes pertaining to the kth particle take the form (see Eqs. (5.27) and (5.31))

$$\frac{d}{dt}(m^{(k)}\gamma^{(k)}\boldsymbol{v}^{(k)}) = \boldsymbol{F}^{(k)}, \tag{5.78}$$

$$\frac{d}{dt}(m^{(k)}c^2\gamma^{(k)}) = \boldsymbol{F}^{(k)}\boldsymbol{v}^{(k)}, \tag{5.79}$$

where $\boldsymbol{F}^{(k)}$ denotes a force acting on the kth particle (summation over k is not performed here!).

If a particle non-interacting with any other particles is considered, $\boldsymbol{F}^{(k)}=0$ and the momentum conservation law $\boldsymbol{p}^{(k)} = m^{(k)}\gamma^{(k)}\boldsymbol{v}^{(k)} = \text{const}$ and the energy conservation law $\mathscr{E}^{(k)} = m^{(k)}\gamma^{(k)}c^2 = \text{const}$ follow directly from Eqs. (5.78) and (5.79). In fact, this circumstance manifests itself in the following relation for an individual particle: $\vec{P}^2 = \boldsymbol{p}^2 - \mathscr{E}^2/c^2 = \text{const}$. In the case of an individual particle which represents a closed system $\vec{P^2}$ remains constant because $\boldsymbol{p}$ and $\mathscr{E}/c$ do individually. Note here once again that $\boldsymbol{p}$ and $i\mathscr{E}/c$ of an individual particle combine to form a 4-vector.

When we deal with a system of n non-interacting material particles, the total momentum $\sum \boldsymbol{p}^{(k)}$ and the total energy of a system $\sum \mathscr{E}^{(k)}$ obviously remain constant because each individual addendum does.

The transformation laws for a total momentum and total energy

$$\boldsymbol{P} = \sum \boldsymbol{p}^{(k)}, \quad \mathscr{E} = \sum \mathscr{E}^{(k)} \tag{5.80}$$

in the case of transition from one IFR to another are evident: a sum of vector components is transformed as a vector component.

The problem of conservation laws in a system of n interacting particles is far more complicated. In conventional classical mechanics the interaction of particles in the case of conservative forces could be described by means of the system's potential function $U = (\boldsymbol{r}^{(1)}, \boldsymbol{r}^{(2)}, \ldots, \boldsymbol{r}^{(n)})$, where $\boldsymbol{r}^{(k)}(t)$ determines the position of the kth particle at the moment t, the positions of all n particles being considered at the same moment of time. In the final analysis a single moment of time can be chosen in classical mechanics due to the assumption of the infinite velocity of interaction propagation.

Since the interaction propagation velocity is finite in relativistic mechanics, the calculation of the force at a given point requires the positions occupied by all particles at some preceding moment to be known. Hence, it is clear that the form which the function U takes in a relativistic case is rather complicated.

If one writes down the expression for the energy of a system of n objects in the form

$$\mathscr{E} = \sum m^{(k)}c^2\gamma^{(k)} \tag{5.81}$$

and for the total momentum

$$\boldsymbol{P} = \sum m^{(k)}\gamma^{(k)}\boldsymbol{v}^{(k)}, \tag{5.82}$$

the following assertion is possible. The quantities $\boldsymbol{P}$ and $i\mathscr{E}/c$ do not form a 4-vector in contrast to what we had for an individual

particle. Besides, these quantities are not constant. The equation $\mathscr{E}$ = const does not hold because in classical mechanics the system's total energy, including its potential energy, remains constant. Eq. (5.81) does not take into account the potential energy, and there is no simple way of incorporating it rigorously. The finite velocity of interaction propagation causes Eq. (5.82) to vary with time. In the final analysis this circumstance clarifies the paradoxycal fact that the quantities $\mathscr{E}$ and $\boldsymbol{P}$ representing the sum of 4-vector components are not 4-vector components themselves. Indeed, in any reference frame where the sums (5.81) and (5.82) are composed, the summands are taken simultaneously in the sense of simultaneity of a given reference frame. When passing over to another inertial frame, one can find the values of moments and energies of individual particles and add them according to the rules of the 4-vector transformation. In a new frame, however, recalculated events will not be simultaneous. In order to find $\mathscr{E}$ and $\boldsymbol{P}$ in a new frame, one has to reduce these sums to simultaneity in this new reference frame. It is this simultaneity recalculation that deprives the quantities $\boldsymbol{P}$ and $\mathscr{E}$ of properties of 4-vector components.

Interacting relativistic systems feature ten integrals of motion: the integral of energy, of momentum, of motion of the centre of inertia, of moment of momentum etc. The approximate appearance of these integrals is presented in the book [16], § 27, for example.

As to the behaviour of integrals of motion in transition from one inertial frame to another, the energy and momentum form a 4-vector, and the integrals of motion of the centre of inertia and of the moment of momentum form an antisymmetric 4-tensor, in the approximation ($\beta^{(k)} = v^{(k)}/c \ll 1$), where the terms $(\beta^{(k)})^2$ are retained. Hence, it is clear that if all these integrals remain constant in one reference frame, they will be constant in any other reference frame.

There is a case in which the conservation laws for momentum and energy can be put down in a simple form:

$$\sum m^{(k)}\gamma^{(k)}\boldsymbol{v}^{(k)} = \sum m^{(k)}\gamma'^{(k)}\boldsymbol{v}'^{(k)}, \tag{5.83}$$

$$\sum m^{(k)}c^2\gamma^{(k)} = \sum m^{(k)}c^2\gamma'^{(k)}. \tag{5.84}$$

These equations are valid only in the case of fast particles which interact weakly (or briefly). Eqs. (5.83) and (5.84) are not valid during an interaction, but they are quite suitable before and after it. In particular, they can be applied to ideal relativistic gas and also to "collisions of microparticles".

Here is an example showing how the conservation laws in a relativistic form are applied in the study of particle "collisions". Let the particle of the mass m_0 strike on a resting particle of the

mass m_1. The total mass of particles generated as a result of the "collision" ("reaction") is equal to M. Reactions between particles are governed not only by the momentum and energy conservation laws, but also by other specific laws of conservation. We shall not take them into consideration here. We shall be assuming, e.g. from experimental data, that a reaction may take place. The momentum and energy conservation laws make it possible to clear up the essential question: what is the minimum energy of a striking particle sufficient to bring about the reaction we are interested in?

"Before" and "after" the reaction the momentum and energy conservation laws are complied with. In four-dimensional terms it means that the 4-vector energy-momentum of a system of particles remains constant. We consider the situation "before" the collision in a laboratory reference frame. Before the collision the particles do not interact, and therefore the energy of the system of particles is equal to $\mathscr{E}_0 + m_1c^2$ and the momentum to $\boldsymbol{p}_0$, where $\mathscr{E}_0$ is the total energy of a striking particle and $\boldsymbol{p}_0$ is its momentum.

The situation after the collision is convenient to consider in the frame of the centre of inertia. In accordance with the momentum conservation law this reference frame moves uniformly and rectilinearly relative to the laboratory one and therefore is also inertial (provided the laboratory frame is inertial). The minimum energy required for the accomplishment of the reaction is realized in the case when all particles generated after the reaction are resting in the frame of the centre of inertia (otherwise their total energy would be higher). Consequently, if the minimum energy is sought for, the energy of a system of particles generated after the collision is equal to Mc^2, and the momentum is equal to zero (in the frame of the centre of inertia). In transitions from one IFR to another the square of a 4-vector energy-momentum is an invariant.

Write out 4-vectors of energy-momentum for a system of particles before and after the collision: $\vec{P}^0(\boldsymbol{p}_0, \mathscr{E}_0/c + m_1c)$, $\vec{P}(0, Mc)$. But the absolute values of these vectors are equal to $\vec{P}^{0^2} = \vec{P}^2$ or $M^2c^2 = (\mathscr{E}_0/c + m_1c)^2 - p_0^2$. Making use of the relations $\mathscr{E}_0^2/c^2 - p_0^2 = m_0^2c^2$, $\mathscr{E}_0 = T_0 + m_0c^2$, we get

$$M^2c^2 = m_0^2c^2 + m_1^2c^2 + 2m_1(T_0 + m_0c^2).$$

Hence, the minimum ("threshold") value of the kinetic energy of a striking particle follows directly:

$$T_0 = \frac{c^2}{2m_1}(M - m_0 - m_1)(M + m_0 + m_1).$$

This formula can be used in the treatment of various reactions. We shall quote three examples.

The production of a π-meson in a collision of two nucleons: $N + N \rightarrow N + N + \pi$. The photoproduction of a π-meson at a nucleon: $N + \gamma \rightarrow N + \pi$. The production of the proton-antiproton pair $(p + \bar{p})$ on a bombardment of a proton (hydrogen)-containing target with protons: $p + p \rightarrow p + p + (p + \bar{p})$.

The interpretation of these reactions has to be based on the fact that the energy conservation law is always observed. In consequence, in these reactions the kinetic energy of initial particles turns (partially) into the rest energy of the particles produced. Speaking of the "production" of mass from kinetic energy is, of course, inaccurate.

CHAPTER 6

THE MAXWELL THEORY IN A RELATIVISTIC FORM

The theory of relativity shows how to consider physical phenomena in any inertial frame of reference. The STR is based on complete equivalence of all inertial frames. Therefore, the basic equations describing physical phenomena in nature must be identical in all inertial frames; of course, in each reference frame they are written in requisite variables, i.e. using length and time scales of a given reference frame.

The basic system of equations describing electromagnetic phenomena was provided by Maxwell. It is remarkable that the system of Maxwell's equations formulated fifty years prior to the advent of the special theory of relativity proved to be covariant with respect to the Lorentz transformation, i.e. it retains its appearance, with the accuracy of variables' designations, under the Lorentz transformation. This signifies that the system of Maxwell's equations retains its appearance in any inertial frame of reference, and the principle of relativity holds automatically.

Thus, the equations of electrodynamics require no modifications in terms of the STR, and it might seem that the theory of relativity cannot introduce anything of importance. It is however far from being the case.

First of all, previous to the advent of the theory of relativity it was not clear in which reference frames the system of Maxwell's equations was valid. The theory of relativity indicated at once that this system of equations fitted any inertial frame of reference. Hence, it was natural to rewrite the system of Maxwell's equations in a four-dimensional form. Such a notation makes it possible to find the transformation equations for the basic quantities of the theory when passing from one IFR to another. Using a four-dimensional notation, we shall also find the inseparable unity of charges and currents, electric and magnetic moments, electric and magnetic fields. Relationship between some other physical quantities will be discovered as well. Such a close relationship between definite physical quantities remained unnoticed until the emergence of a relativistic approach to electromagnetic phenomena.

As to the transformation of electric and magnetic field components on transition from one inertial frame to another, it can be carried out systematically only in terms of the theory of relativity. It is the theory of relativity that indicates a four-dimensional antisymmetric tensor to be used to describe an electromagnetic field.

§ 6.1. The three-dimensional system of Maxwell's equations. A 4-potential and 4-current *. The Maxwell theory represents a macroscopic theory of electromagnetic field. In accordance with this theory an electromagnetic field in an arbitrary medium is described by the four vectors: the electric field strength $\boldsymbol{E}$, magnetic field strength $\boldsymbol{H}$, electric field induction $\boldsymbol{D}$ and magnetic field induction $\boldsymbol{B}$.

In a uniform isotropic medium the number of field vectors needed to describe electromagnetic phenomena is reduced to two since field vectors turn out to be proportional to each other:

$$\boldsymbol{D} = \varepsilon \boldsymbol{E}, \quad \boldsymbol{B} = \mu \boldsymbol{H}. \tag{6.1}$$

The constant coefficients ε and μ are called a dielectric permittivity and magnetic permeability respectively. Vacuum is described as a uniform isotropic medium possessing the definite values of ε and μ which are customarily denoted by ε_0 and μ_0 and referred to as an electric and a magnetic constant respectively.

According to the Maxwell theory the field vectors obey the two principal equations:

$$\left.\begin{aligned} \operatorname{rot} \boldsymbol{H} &= \boldsymbol{j} + \dot{\boldsymbol{D}}, \\ \operatorname{rot} \boldsymbol{E} &= -\dot{\boldsymbol{B}}; \end{aligned} \quad \text{(a)} \quad \right| \quad \begin{aligned} \operatorname{rot} \boldsymbol{H} &= \boldsymbol{j} + \varepsilon\dot{\boldsymbol{E}}, \\ \operatorname{rot} \boldsymbol{E} &= -\mu\dot{\boldsymbol{H}}; \end{aligned} \quad \text{(b)} \tag{6.2}$$

the equations for an arbitrary medium are written on the left-hand side; the same equations for a uniform and isotropic medium are seen on the right-hand side.

In the Maxwell theory the average values of an electric and a magnetic field (relative to "actual" microscopic fields) are defined by the vectors $\boldsymbol{E}$ and $\boldsymbol{B}$. In the general case the vectors $\boldsymbol{D}$ and $\boldsymbol{H}$ are related to the average fields by the following equations:

$$\boldsymbol{D} = \varepsilon_0 \boldsymbol{E} + \boldsymbol{P}, \quad \boldsymbol{B} = \mu_0 (\boldsymbol{H} + \boldsymbol{M}), \tag{6.3}$$

where two more vectors are introduced: the polarization vector $\boldsymbol{P}$ and magnetization vector $\boldsymbol{M}$.

The charge conservation law is assumed to be valid in the Maxwell theory; in the case of a continuous charge distribution it is

* All equations of electrodynamics are written in this chapter by means of the SI units. For the reader's convenience the principal formulae are also written out in the Gaussian system of units in Appendix II.

written down as the continuity equation:

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \boldsymbol{j} = 0. \tag{6.4}$$

Here ρ is a charge density and $\boldsymbol{j} = \rho \boldsymbol{v}$ a current density.

From Eqs. (6.2) and (6.4) the two equations follow which are easily incorporated in (6.2) and (6.4):

$$\left.\begin{matrix} \operatorname{div} \boldsymbol{D} = \rho, \\ \operatorname{div} \boldsymbol{B} = 0; \end{matrix} \quad \text{(a)} \quad \right| \quad \begin{matrix} \operatorname{div} \boldsymbol{E} = \rho/\varepsilon, \\ \operatorname{div} \boldsymbol{H} = 0. \end{matrix} \quad \text{(b)} \tag{6.5}$$

The force density for an electromagnetic field acting on free charges and currents is assumed to be equal to

$$\boldsymbol{f} = \rho \{\boldsymbol{E} + [\boldsymbol{v}\boldsymbol{B}]\}. \tag{6.6}$$

It is referred to as the Lorentz force. From this expression it is seen once again that $\boldsymbol{E}$ and $\boldsymbol{B}$ are average macroscopic fields.

The system of Maxwell's equations can be written down not only by means of field vectors but also via the scalar and vector potentials φ and $\boldsymbol{A}$. We shall consider the case of a uniform and isotropic medium and combine the potentials φ and $\boldsymbol{A}$ with the fields $\boldsymbol{E}$ and $\boldsymbol{B}$ by means of the relations

$$\boldsymbol{E} = -\nabla\varphi - \dot{\boldsymbol{A}}, \quad \boldsymbol{B} = \operatorname{rot} \boldsymbol{A}. \tag{6.7}$$

Having substituted these expressions into Eq. (6.2b) and having additionally imposed the Lorentz condition

$$\operatorname{div} \boldsymbol{A} + \frac{1}{v^2} \dot{\varphi} = 0 \tag{6.8}$$

on potentials (one can show this condition to be easily met), we obtain the equations which hold in the case of the potentials φ and $\boldsymbol{A}$:

$$\Box \varphi = -\frac{\rho}{\varepsilon}, \quad \Box \boldsymbol{A} = -\mu \boldsymbol{j}, \tag{6.9}$$

where

$$\Box = \Delta - \frac{1}{v^2} \frac{\partial^2}{\partial t^2}, \quad v = \frac{1}{\sqrt{\varepsilon\mu}}. \tag{6.10}$$

In Eqs. (6.9) it is assumed that $\rho = \rho(\boldsymbol{r}, t)$, $\boldsymbol{j} = \boldsymbol{j}(\boldsymbol{r}, t)$, i.e. the current and charge densities are the assigned functions of coordinates and time. Eqs. (6.7) and (6.9) are equivalent to (6.2).

Next, we shall impart a four-dimensional meaning to the quantities involved in the Maxwell equations and rewrite the Maxwell equations themselves in a four-dimensional form. However, we shall proceed gradually, and the Maxwell equations (6.2) and

(6.4) will be written in a four-dimensional form only in § 6.7. Now we begin with composing four-dimensional quantities from the potentials φ, $\boldsymbol{A}$ and the densities ρ, $\rho\boldsymbol{v}$.

Eqs. (6.9) are differential equations of the same kind, the d'Alembertian equations. Accordingly, they can be written at once as a single four-dimensional equation, provided that the two 4-vectors are introduced the 4-potential vector $\vec{\Phi}$ and the 4-current density vector $\vec{s}$.

For some time we shall be simultaneously writing out definitions and relations in the real and imaginary forms, just the way we did it in mechanics. So, let us define the 4-potential vector $\vec{\Phi}$ as follows:

$$\vec{\Phi}\begin{Bmatrix} \Phi_1 & \Phi_2 & \Phi_3 & \Phi_4 \\ A_x & A_y & A_z & (i/c)\varphi \end{Bmatrix}, \quad \text{(a)} \quad \Bigg| \quad \vec{\Phi}\begin{Bmatrix} \Phi^0 & \Phi^1 & \Phi^2 & \Phi^3 \\ \varphi/c & A_x & A_y & A_z \end{Bmatrix}, \quad \text{(b)} \tag{6.11}$$

and 4-current vector

$$\vec{s}\begin{Bmatrix} s_1 & s_2 & s_3 & s_4 \\ j_x & j_y & j_z & ic\rho \end{Bmatrix}, \quad \text{(a)} \quad \Bigg| \quad \vec{s}\begin{Bmatrix} s^0 & s^1 & s^2 & s^3 \\ c\rho & j_x & j_y & j_z \end{Bmatrix}. \quad \text{(b)} \tag{6.12}$$

Recall the definition of the 4-radius vector:

$$\vec{R}\begin{Bmatrix} x_1 & x_2 & x_3 & x_4 \\ x & y & z & ict \end{Bmatrix}, \quad \text{(a)} \quad \Bigg| \quad \vec{R}\begin{Bmatrix} x^0 & x^1 & x^2 & x^3 \\ \tau = ct & x & y & z \end{Bmatrix}. \quad \text{(b)}$$

The 4-vector components $\vec{\Phi}$, $\vec{s}$, $\vec{R}$ are written in two parallel lines, in a conventional and a symmetric notation respectively; the necessary component values are immediately found from the comparison of these lines. Having determined the 4-potential and 4-current density, one can rewrite Eq. (6.9) for vacuum (i.e. with $\varepsilon = \varepsilon_0$ and $\mu = \mu_0$, where $c^2 = 1/(\varepsilon_0\mu_0)$) as a unified formula:

$$\Box\,\Phi_k = -\mu_0 s_k \ (k = 1, 2, 3, 4), \quad \text{(a)} \quad \Big| \quad \Box\,\Phi^k = -\mu_0 s^k \ (k = 0, 1, 2, 3). \quad \text{(b)} \tag{6.13}$$

Obviously, the three equations of (6.13a) at $k = 1, 2, 3$ coincide with the three equations of (6.9) in the case of vacuum. At $k = 4$ Eq. (6.13a) gives $\Box \frac{i}{c}\varphi = -\mu_0 ic\rho$, and since $c^2 = 1/(\varepsilon_0\mu_0)$, we obtain Eq. (6.9) again.

We suggest that the reader make sure that Eqs. (6.13b) also coincide with Eqs. (6.9) in vacuum.

In the case of vacuum the Lorentz condition (Eq. (6.8)) and the charge conservation law (Eq. (6.4)) can be written via the

4-divergence of the vectors $\vec{\Phi}$ and $\vec{s}$. Indeed, for example,

$$\operatorname{div}\vec{\Phi}=\frac{\partial\Phi_i}{\partial x_i}=\frac{\partial\Phi_1}{\partial x_1}+\frac{\partial\Phi_2}{\partial x_2}+\frac{\partial\Phi_3}{\partial x_3}+\frac{\partial\Phi_4}{\partial x_4}=$$
$$=\frac{\partial A_x}{\partial x}+\frac{\partial A_y}{\partial y}+\frac{\partial A_z}{\partial z}+\frac{\partial (i\varphi/c)}{\partial (ict)}=\operatorname{div}\boldsymbol{A}\frac{1}{c^2}\dot{\varphi},$$
$$\operatorname{div}\vec{s}=\frac{\partial s_i}{\partial x_i}=\frac{\partial j_x}{\partial x}+\frac{\partial j_y}{\partial y}+\frac{\partial j_z}{\partial z}+\frac{\partial (ic\rho)}{\partial (ict)}=\operatorname{div}\boldsymbol{j}+\frac{\partial\rho}{\partial t}.$$

Consequently, the Lorentz condition and the charge conservation law in vacuum have the form div $\vec{\Phi}=0$, div $\vec{s}=0$, when presented in a four-dimensional notation. Surely, the same results are obtained on the basis of the usual three-dimensional approach. The conclusions just drawn are very significant. As it is shown in Appendix I, § 4, the 4-divergence is the Lorentz transformation invariant. As to Eqs. (6.13), these 4-vector relations are valid in any inertial frame of reference in vacuum. Thus, the equations for potentials, the Lorentz condition and the charge conservation law can be so rewritten that it would become evident at once that they retain their appearance in any inertial frame of reference. The covariant notation of equations for potentials in the case of a refractive medium ($\varepsilon\neq\varepsilon_0$ and $\mu\neq\mu_0$) will be discussed below (see §§ 6.14, 6.15).

§ 6.2. The transformation of a 4-potential and 4-current. The very fact that we have managed to compose the 4-vectors $\vec{\Phi}$ and $\vec{s}$ makes it possible to write down immediately the transformation equations for components of these vectors. We shall write these equations here in both the real and the imaginary form (cf. Eq. (4.10a, b)):

$$\Phi_1=\Gamma(\Phi_1'-iB\Phi_4'),\quad \Phi_2=\Phi_2',\quad \Phi_3=\Phi_3',\quad \Phi_4=\Gamma(\Phi_4'+iB\Phi_1'); \tag{6.14a}$$

$$\Phi^0=\Gamma(\Phi^{0'}+B\Phi^{1'}),\quad \Phi^1=\Gamma(\Phi^{1'}+B\Phi^{0'}),\quad \Phi^2=\Phi^{2'},\quad \Phi^3=\Phi^{3'}; \tag{6.14b}$$

$$s_1=\Gamma(s_1'-iBs_4'),\quad s_2=s_2',\quad s_3=s_3',\quad s_4=\Gamma(s_4'+iBs_1'); \tag{6.15a}$$

$$s^0=\Gamma(s^{0'}+Bs^{1'}),\quad s^1=\Gamma(s^{1'}+Bs^{0'}),\quad s^2=s^{2'},\quad s^3=s^{3'}. \tag{6.15b}$$

Let us look more carefully into the current density transformation. A 4-current comprises a current density and a charge density. It is quite natural that a current and a charge density are combined into a single 4-vector. Dealing with reference frames moving relative to one another, one should bear in mind that a charge may be at rest only in one ("proper") reference frame. In all other IFRs the charge moves, being in terms of these frames

not only a charge but also a current. We see, therefore, how easy is the transition from a motionless charge (electrostatics) to a moving one (current): it is just a transition from the charge's proper reference frame to any other IFR. When a current originates due to charges displaced together with a moving medium or objects, it is called a convection current. It is just a *convection current* that turns up on transition from a proper frame to an arbitrary IFR.

The formula $j = \rho v$ contains the density of charges moving at the velocity v. Otherwise, misunderstandings may arise. For example, a current flows in metals, even though $\rho = 0$. Indeed, the total charge density in metals comprises those of ions and free electrons and is equal to zero: $\rho = \rho_+ + \rho_- = 0$. But certainly, a current can flow provided there is a regular motion of electrons: $j = \rho_+ v_+ + \rho_- v_- = \rho_- v_-$, because the velocity of a regular motion of ions is equal to zero.

From Eqs. (6.15) we obtain directly a convection current originating on transition from the charge's "proper" frame. Thus, let the frame K' be characterized by a given charge density ρ' and the absence of a current ($j' = 0$). Consequently, in the frame K' the 4-current density has the components $\vec{s}'$ $(0, 0, 0, ic\rho' = ic\rho_0)$, i.e. $s_1' = s_2' = s_3' = 0$, $s_4' = ic\rho_0$. Then in accordance with Eq. (6.15a), for example, in the frame K

$$s_1 = \Gamma(-i\mathrm{B}ic\rho_0) = \Gamma V\rho_0, \quad s_2 = s_3 = 0, \quad s_4 = \Gamma ic\rho_0. \tag{6.16}$$

In the developed form the last relation of Eq. (6.16) has the following appearance:

$$s_4 \equiv ic\rho = \frac{ic\rho_0}{\sqrt{1 - V^2/c^2}}.$$

This way we obtain the charge density transformation law on transition from the charge's "proper" frame, in which charges are at rest, to the frame relative to which the charges move at the velocity V:

$$\rho = \frac{\rho_0}{\sqrt{1 - V^2/c^2}}. \tag{6.17}$$

The first relation of (6.16) yields a current density

$$s_1 = j_x = \frac{\rho_0 V}{\sqrt{1 - V^2/c^2}} = \rho V. \tag{6.18}$$

As we have mentioned, a current associated with the motion of a charged medium or charged object is referred to as a convection current.

The meaning of Eq. (6.18) is very simple. The velocity of a charge resting in the frame K' is equal to V with respect to the

frame K (that is the velocity of the frame K'; the same follows from the relativistic velocity transformation formula (3.27)). Therefore, Eq. (6.18) represents a convection current. As to the charge density transformation (see Eq. (6.17)), it is associated with the volume change (since a density is a charge of a volume unit). Since a volume transforms according to the law

$$d\mathscr{V} = d\mathscr{V}_0 \sqrt{1 - V^2/c^2},$$

and we consider the same physical volume containing the same charge de, then

$$\rho_0 = \frac{de}{d\mathscr{V}_0}, \text{ and } \rho = \frac{de}{d\mathscr{V}} = \frac{1}{\sqrt{1 - V^2/c^2}} \frac{de}{d\mathscr{V}_0} = \frac{\rho_0}{\sqrt{1 - V^2/c^2}} = \rho_0 \Gamma.$$

Of course, the total charge in a given volume remains constant in any reference frame:

$$\rho_0 d\mathscr{V}_0 = \rho \, d\mathscr{V}. \tag{6.19}$$

Eq. (6.19) expresses the invariance of a charge confined in a given volume. Using Eq. (6.17), the 4-vector $\vec{s}$ can be expressed otherwise. Let us consider a small volume element of a moving medium in the frame K. Then in the reference frame K^0 co-moving with this element the velocity of the element $v = 0$ and $\rho = \rho_0$. In the frame K the density $\rho = \rho_0\gamma$ and, consequently, $j = \rho v = \rho_0\gamma v$. Then $\vec{s}(\rho_0\gamma v, \rho_0 ic\gamma) = \rho_0\vec{V}$, where $\vec{V}$ is the 4-velocity of an element of a medium. Thus,

$$\vec{s} = \rho_0\vec{V}, \quad s_i = \rho_0 u_i. \tag{6.20}$$

For $\vec{s}^2$ we have the following form:

$$\vec{s}^2 = \rho_0^2\gamma^2 (v^2 - c^2) = -\rho_0^2/c^2. \tag{6.21}$$

The 4-vector $\vec{s}$ is a time-like vector because a charge velocity v is always less than c.

If in the frame K' there is an uncharged conductor through which a current flows, i.e.

$$\vec{s}_0 (j_{x0}, \; j_{y0}, \; j_{z0}, \; ic\rho_0 = 0), \tag{6.22}$$

in the frame K', a certain charge density ρ is observed in K. Indeed, according to Eq. (6.15),

$$s_1 = \Gamma j_{x0}, \quad s_2 = j_{y0}, \quad s_3 = j_{z0}, \quad s_4 \equiv ic\rho = \Gamma i\mathrm{B} j_{x0}. \tag{6.23}$$

The first three formulae of (6.23) define the current magnitude in the frame K, the last one defines the charge density in the

frame K:

$$\rho = \Gamma \frac{B}{c} j_{x0}. \quad (6.24)$$

Consequently, an observer in K will find the density ρ, even though the charge density in the frame K' is equal to zero. This result can easily be interpreted in geometric terms. Let the conductor be at rest in the frame K; the ions of the conductor are motionless and the electrons move at some average velocity v. In

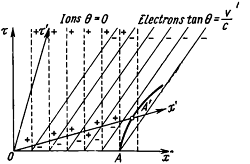


Fig. 6.1. The Minkowski diagram illustrating the emergence of a charge density in the frame K' in a current conductor which has a zero charge density in the frame K. The total charge density caused by the ions and electrons in a conductor is equal to zero. Current is generated due to the motion of the electrons, with the ions being motionless The world lines of the ions are depicted by dotted lines, and the world lines of the electrons by slanted continuous lines. Besides the reference frame K (with the x, τ axes), also shown is the reference frame K' (with the x', τ' axes); a scale hyperbola cutting unitary segments on the x and x' axes is also drawn in the figure. Inasmuch as a charge density must be determined simultaneously at all points, it has to be determined at all points of the object on the x or x' axis respectively. It is seen that if a unitary segment in the frame K contains an equal number of ions and electrons, a unitary segment in the frame K' will contain more ions than electrons. This fact signifies the emergence of a positive charge density in the frame K'

the frame K the world lines of the ions are straight lines parallel to the τ axis while the world lines of the electrons are straight lines forming a certain angle $\theta = \arctan(v/c)$ with the τ axis.

Fig. 6.1 shows the reference frames $K(x, \tau)$ and $K'(x', \tau')$, the world lines of ions (dotted lines), and the world lines of electrons (thin continuous lines inclined to the τ axis at an angle θ). Inasmuch as metal is neutral on the average, each segment of the conductor must emit an equal number of world lines of ions and electrons. The charge density should be measured simultaneously in each reference frame. In the frame K it is determined by the number of world lines of ions and electrons crossing a unit of length in this frame. For example, the charge density is defined

by the number of world lines of ions (taken with the sign "+") and the number of world lines of electrons (taken with the sign "—") going through the segment OA. A scale hyperbola cuts the unit segments on the x and x' axes. It is these unit segmentsthat the charges must be related to. But in the frame K' the charge density must be calculated for the whole conductor simultaneously. In the frame K' simultaneous events are located on straight lines parallel to the x' axis, and on the x' axis itself, in particular. It is seen, however, from Fig. 6.1 that the unit segment OA' accommodates more positive charges than negative ones. Accordingly, the conductor will turn up to be charged positively in the frame K' although it is neutral in the frame K. Surely, if one considers a closed-type conductor in the frame K', its total charge will remain equal to zero, but an electric dipole moment, not found in the frame K, will be observed in the frame K' (see § 6.9 and Fig. 6.4).

§ **6.3. An electromagnetic field tensor.** In electrodynamics the electric field strength $\boldsymbol{E}$ and magnetic induction $\boldsymbol{B}$ are conveniently expressed via the vector and scalar potentials $\boldsymbol{A}$ and φ as follows *:

$$\boldsymbol{B} = \operatorname{rot} \boldsymbol{A}, \quad \boldsymbol{E} = -\operatorname{grad} \varphi - \partial \boldsymbol{A}/\partial t. \tag{6.25}$$

Let us rewrite these equations using 4-potential components; for the present, we shall write out the relations of the complex 4-space:

$$B_x \equiv B_1 = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = \frac{\partial \Phi_3}{\partial x_2} - \frac{\partial \Phi_2}{\partial x_3}. \tag{6.26}$$

$$E_x \equiv E_1 = -\frac{\partial \varphi}{\partial x} - \frac{\partial A_x}{\partial t} = -\frac{c}{i}\frac{\partial \Phi_4}{\partial x_1} - \frac{\partial \Phi_1}{\partial x_4} ic = ic\left(\frac{\partial \Phi_4}{\partial x_1} - \frac{\partial \Phi_1}{\partial x_4}\right). \tag{6.27}$$

The last terms of Eqs. (6.26) and (6.27) are written on the basis of the definition of 4-potential components. Similarly, using the $\vec{\Phi}$ components, one can write down the remaining components of the vectors $\boldsymbol{E}$ and $\boldsymbol{B}$ as well. We shall obtain equations similar to Eqs. (6.26) and (6.27) from which it follows that all components of the vectors $\boldsymbol{E}$ and $\boldsymbol{B}$ can be expressed via certain combinations of derivatives of the 4-vector $\vec{\Phi}$ components with respect

* In previous chapters we utilized the letter B to denote the ratio of the velocity of a coordinate system to that of light Here we have to introduce a magnetic induction vector $\boldsymbol{B}$ and its projections B_x, B_y, B_z To eliminate blunders, we shall not be using the designation B = V/c in this chapter, apart from the cases when misunderstandings are ruled out

to four-dimensional coordinates. These combinations form the antisymmetric 4-tensor of the second rank *:

$$F_{ik} = c\left(\frac{\partial \Phi_k}{\partial x_i} - \frac{\partial \Phi_i}{\partial x_k}\right) \quad (i,\ k = 1,\ 2,\ 3,\ 4). \tag{6.28}$$

Prior to discussing mathematical features and the meaning of Eq. (6.28) we should examine the same transition in the real 4-space. As we have indicated, one has to differentiate between covariant and contravariant components of vectors and tensors in this case. The electromagnetic field tensor (Eq. (6.28)) is expediently written in covariant components. Then, if the vector $\vec{\Phi}$ has the contravariant components $(\Phi^0, \boldsymbol{A})$, its covariant components will be $(\Phi^0, -\boldsymbol{A})$. The differentiation with respect to contravariant coordinates leads to covariant components again (see Appendix I, § 8). Thus, Eq. (6.26) only changes the sign while Eq. (6.27) is written as

$$E_x \equiv E_1 = -c\frac{\partial \Phi_0}{\partial x^1} + c\frac{\partial \Phi_1}{\partial x^0} = c\left(\frac{\partial \Phi_1}{\partial x^0} - \frac{\partial \Phi_0}{\partial x^1}\right), \tag{6.27'}$$

taking into account that $\varphi = c\Phi_0$, $A_x = -\Phi_1$, $t = x^0/c$, $x = x^1$. Thus, in the case of the real 4-space we introduce the covariant antisymmetric 4-tensor of the second rank

$$F_{ik} = c\left(\frac{\partial \Phi_k}{\partial x^i} - \frac{\partial \Phi_i}{\partial x^k}\right), \tag{6.28'}$$

coinciding in its appearance with Eq. (6.28). The purpose of the introduction of Eq. (6.28) or (6.28′) is very remarkable. The two Maxwellian field vectors $\boldsymbol{E}$ and $\boldsymbol{B}$ can be expressed in the 4-space uniquely through a certain combination of space-time derivatives of the 4-vector potential $\vec{\Phi}$. The behaviour of the quantities F_{ik} on transition from one IFR to another is most significant for us. Their transformation, however, is very easy to find: the quantities F_{ik} form a tensor since one can readily make sure (see Appendix I, § 3) that derivatives of 4-vector components with respect to coordinates transform in accordance with the tensor transformation rule. If the indices i and k in Eqs. (6.28) and (6.28′) acquire independently all values from 1 to 4 (and from 0 to 3 respectively), we obtain 16 values of F_{ik} (four of which are equal to zero) expressed through components of $\boldsymbol{E}$ and $\boldsymbol{B}$. Let us write

* Just as in most books on relativistic electrodynamics, we use the subindices i and k despite the fact that the imaginary unity, also designated by i, keeps recurring alongside over and over again. We hope that this will not lead to a misunderstanding.

down these components in the form of matrices:

$$F_{ik}=\begin{pmatrix} 0 & cB_z & -cB_y & -iE_x \\ -cB_z & 0 & cB_x & -iE_y \\ cB_y & -cB_x & 0 & -iE_z \\ iE_x & iE_y & iE_z & 0 \end{pmatrix}, \quad \text{(a)}$$

$$F_{ik}=\begin{pmatrix} 0 & cE_x & cE_y & cE_z \\ -cE_x & 0 & -cB_z & cB_y \\ -cE_y & cB_z & 0 & -cB_x \\ -cE_z & -cB_y & cB_x & 0 \end{pmatrix}. \quad \text{(b)} \qquad (6.29)$$

We see that the components of an electric field strength and magnetic induction are the components of one 4-tensor of an electromagnetic field. As usual, in the designation F_{ik} the first sub-index i denotes the line and the second k the column of the matrix F_{ik}.

Frequently the tensor (6.29a) is abbreviated as $\mathfrak{F}=(c\boldsymbol{B}, -i\boldsymbol{E})$ * and the tensor (6.29b) as $\mathfrak{F}=(\boldsymbol{E}, c\boldsymbol{B})$, assuming the components of the vectors $\boldsymbol{E}$ and $\boldsymbol{B}$ are arranged as in (6.29a) and (6.29b) respectively.

We have obtained the result which is dissimilar to the usual three-dimensional case. The Maxwell theory deals customarily with vector fields. In fact, the vectors $\boldsymbol{E}$ and $\boldsymbol{B}$ behave as 3-vectors as far as the transformation of a coordinate system, that is the rotation of coordinate axes, is concerned. As soon as we pass over to reference frames moving relative to one another, the situation changes drastically. In 4-space $\boldsymbol{E}$ and $\boldsymbol{B}$ are no longer vectors, not even four-dimensional ones. Although the vectors $\boldsymbol{E}$ and $\boldsymbol{B}$ are expressed via components of a four-dimensional potential, there are no values to be added to the three-dimensional vectors $\boldsymbol{E}$ and $\boldsymbol{B}$ in order to make them become 4-vectors. In 4-space an electromagnetic field is a single quantity of a more complicated mathematical nature than a 4-vector. The fields $\boldsymbol{E}$ and $\boldsymbol{B}$ have merged into a single 4-tensor which is referred to as an *electromagnetic field tensor.*

The emergence of a single 4-tensor instead of two three-dimensional vectors describing an electromagnetic field has a clear-cut physical meaning. Electric and magnetic fields are intertwined so inseparably that the "appearance" or "disappearance" of one of the fields is determined by the choice of a reference frame. For example, a "pure" electric field generated by a charge occurs under very specific circumstances when the charge is considered

* The Gothic letter $\mathfrak{F}$ denotes 4-force components in Chapter 5 In this chapter the same letter is used exclusively for the designation of a tensor.

in the frame in which it is at rest. In any other inertial frame, however, this charge moves and consequently generates an electric current producing a magnetic field. On the other hand, we have seen that although in a certain reference frame a current-carrying conductor appears neutral, in other inertial frames of reference it appears charged, and consequently, an electric field is due to occur in these frames.

Thus, it is sufficient, for example, to have only an electric field in the frame K for a magnetic field to appear in any other frame K'. If in the frame K there is only a magnetic field, both magnetic and electric fields will appear in any other frame K'. Had we attempted to treat the fields $\boldsymbol{E}$ and $\boldsymbol{B}$ as vectors, that physical fact could not have been expressed in mathematical terms. As we have mentioned, there are no values to make the three-dimensional vectors of an electromagnetic field become 4-vectors. Moreover, had each of the vectors $\boldsymbol{E}$ and $\boldsymbol{B}$ been contained in "its respective" 4-vector, the Lorentz transformation would have necessitated each of these vectors in a "new" frame to be expressed through components of "its respective" vector in an "old" frame. In such a way, the vectors $\boldsymbol{E}$ and $\boldsymbol{B}$ would turn out to be unrelated. Practice, however, indicates an intimate connection between electric and magnetic fields, that is between the vectors $\boldsymbol{E}$ and $\boldsymbol{B}$.

Two three-dimensional vectors possess six independent components. An antisymmetric 4-tensor of the second rank possesses exactly six independent components. We have found (see Eq. (6.29a, b)) that the fields $\boldsymbol{E}$ and $\boldsymbol{B}$ form an antisymmetric 4-tensor, an electromagnetic field tensor. Inasmuch as any component of a tensor in a new reference frame is a linear combination of all components of that tensor in an old reference frame, a transition from one reference frame to another may result in the appearance of an electric field due to a magnetic field observed in another frame, and vice versa. In a certain sense, an electromagnetic field is a closed formation: if some inertial frame has no electric or magnetic fields, an electromagnetic field will not appear in any other inertial frame. We shall deal with the transformation of electromagnetic field components in the next section while here we shall examine briefly an electromagnetic field in matter.

To describe a field in matter, one has to introduce, in addition to the average fields $\boldsymbol{E}$ and $\boldsymbol{B}$, two more vectors. These can be either the electric induction vector $\boldsymbol{D}$ and the magnetic field strength $\boldsymbol{H}$, or the electric polarization vector $\boldsymbol{P}$ and the magnetization vector $\boldsymbol{M}$. These four vectors are interrelated as follows:

$$\boldsymbol{D} = \varepsilon_0 \boldsymbol{E} + \boldsymbol{P}, \quad \boldsymbol{H} = \boldsymbol{B}/\mu_0 - \boldsymbol{M}. \tag{6.30}$$

The vectors $\boldsymbol{H}$ and $\boldsymbol{D}$ form a special tensor whose components are customarily denoted by f_{ik} while the tensor itself is abbreviated

as $\mathfrak{f} = (\boldsymbol{H}, -ic\boldsymbol{D})$. This tensor is derived from (6.29a) by substituting $\boldsymbol{H}$ components for $c\boldsymbol{B}$ components and $-ic\boldsymbol{D}$ components for $-i\boldsymbol{E}$ ones. Here is the matrix of this tensor components:

$$f_{ik}=\begin{pmatrix} 0 & H_z & -H_y & -icD_x \\ -H_z & 0 & H_x & -icD_y \\ H_y & -H_x & 0 & -icD_z \\ icD_x & icD_y & icD_z & 0 \end{pmatrix}. \tag{6.31}$$

In addition to the tensor $\mathfrak{f}$, it is also useful to introduce a tensor of electric and magnetic moments of matter whose definition follows easily from Eq. (6.30):

$$m_{ik}=\sqrt{\frac{\varepsilon_0}{\mu_0}}\,F_{ik}-f_{ik}. \tag{6.32}$$

It is abbreviated as follows: $\mathfrak{M} = (\boldsymbol{M}, ic\boldsymbol{P})$. Written in full it is

$$m_{ik}=\begin{pmatrix} 0 & M_z & -M_y & icP_x \\ -M_z & 0 & M_x & icP_y \\ M_y & -M_x & 0 & icP_z \\ -icP_x & -icP_y & -icP_z & 0 \end{pmatrix}. \tag{6.33}$$

It is the very fact of the origination of tensors (6.29), (6.31) and (6.33) that points to a close pairwise connection between the quantities $\boldsymbol{E}$, $\boldsymbol{B}$; $\boldsymbol{H}$, $\boldsymbol{D}$ and $\boldsymbol{M}$, $\boldsymbol{P}$. We have written out here the tensors f_{ik} and m_{ik} only for a 4-complex space; the corresponding expressions for a real 4-space can be easily derived by the reader himself.

§ 6.4. The transformation of electric and magnetic field components. The four-dimensional approach is particularly convenient because as soon as the mathematical nature of one or another physical quantity is established (a scalar, 4-vector, 4-tensor), the problem of its transformation on transition from one IFR to another is solved automatically. In mechanics we dealt with 4-vectors. As we have established, components of the fields $\boldsymbol{E}$ and $\boldsymbol{B}$, $\boldsymbol{H}$ and $\boldsymbol{D}$, and $\boldsymbol{M}$ and $\boldsymbol{P}$ are components of tensors (6.29a, b), (6.31) and (6.33) respectively.

Consequently, components of three-dimensional vectors are transformed according to the role of tensor component transformation. For example, the components F_{ik} in a 4-complex space are transformed as follows:

$$F_{ik}=\alpha_{im}\alpha_{kl}F'_{ml}, \tag{6.34}$$

where α_{im} are components of the Lorentz transformation matrix (2.41a) while components F_{ik} are defined by Eq. (6.29a). In order

to transform components of matrix (6.29b), one has to make use of the Lorentz transformation matrix in the form (2.41b).

Here we ought to make a purely methodological remark. Lecturers often avoid using tensors trying not to complicate a lecture course. Indeed, explaining the meaning of tensors and their properties in the course of half an hour is a difficult task. One cannot, however, disregard the fact that an electromagnetic field is a tensor. The old question arises again: "shall we call a cat a cat" from the very beginning? Of course, it is not so much the matter of a name, as the transformation equation (6.34). Apparently this equation can and should be obtained in the simplest way possible. For example, it is easily derived as follows: we see from Eq. (6.28) that the quantities F_{ik} are linear combinations of 4-vector component derivatives with respect to 4-coordinates; the transformation of vector component derivatives due to the coordinate transformation follows from the simple analysis (see Appendix I, § 3).

To memorize the transformation rules for tensor components, one should keep in mind that they transform as a product of corresponding vector components. One way or another, we get Eq. (6.34). And here it is the right time to call a tensor a tensor, having disclosed, rather one-sidedly, of course, the meaning of tensor components by means of vector component derivatives with respect to coordinates.

We shall give an example of how the field transformation equation can be obtained for the case $B_z = F_{12}/c$. According to Eq. (6.34) the transformation equation for F_{12} has the following form:

$$F_{12} = \alpha_{1m}\alpha_{2l}F'_{ml}. \tag{6.35}$$

Recall that the summation is carried out here over the two independent pairs of the indices m and l, each of which runs from 1 to 4. This way, Eq. (6.35) involves the sum of sixteen terms, each being a product of two α_{ik} and one of F_{ik} components. We urge the readers who come across such equations for the first time to write out (once in a lifetime) all sixteen terms. Here is the easiest way to do this. First, we develop the sum with m taking the values 1, 2, 3, 4. As before, the index l denotes summation. This way we obtain a sum consisting of four terms in which the index m is eliminated. Then we perform the summation over l in each of these four terms. As a result all sixteen terms will be written out. Then one should substitute α_{ik} from the Lorentz matrix (see Eq. (2.41a)) and components F_{ik} from Eq. (6.29a) into these terms. One can see at once that most terms of the sum (6.35) are equal to zero. Because of this the summation in Eq. (6.35) can be much simpler. Indeed, the quantities α_{1m}, with m running from 1 to 4, constitute the elements of the first line of the Lorentz matrix (see

Eq. (2.41a)) while the second line is made up of the quantities α_{2l} with $l = 1, 2, 3, 4$. But the first line of the matrix has only two elements, α_{11} and α_{14}, which differ from zero. Consequently, one must consider only the values of m equal to 1 and 4. The second line has only one element differing from zero, $\alpha_{22} = 1$. Consequently, one has to take only $l = 2$ and to rewrite Eq. (6.35) as

$$F_{12} \equiv cB_z = \alpha_{22}\alpha_{1m}F'_{m2} = \alpha_{1m}F'_{m2} = \alpha_{11}F'_{12} + \alpha_{14}F'_{42} =$$
$$= \Gamma\left\{cB'_z - i\frac{V}{c}(iE'_y)\right\} = \frac{cB'_z + \frac{V}{c}E'_y}{\sqrt{1 - V^2/c^2}}.$$

Intercomparing the second and the last equation in this chain of equations and dividing them by c, we get

$$B_z = \frac{B'_z + \frac{V}{c^2}E'_y}{\sqrt{1 - V^2/c^2}} = \Gamma\left(B'_z + \frac{V}{c^2}E'_y\right).$$

In much the same way we obtain the transformation equations for the other components:

$$\begin{aligned} &E_x = E'_x, \quad E_y = \Gamma(E'_y + VB'_z), \qquad E_z = \Gamma(E'_z - VB'_y); \\ &B_x = B'_x, \quad B_y = \Gamma\left(B'_y - \frac{V}{c^2}E'_z\right), \quad B_z = \Gamma\left(B'_z + \frac{V}{c^2}E'_y\right). \end{aligned} \tag{6.36}$$

Let us write out the transformation equations for $\boldsymbol{D}$ and $\boldsymbol{H}$ to be used later on:

$$\begin{aligned} &D_x = D'_x, \quad D_y = \Gamma\left(D'_y + \frac{V}{c^2}H'_z\right), \quad D_z = \Gamma\left(D'_z - \frac{V}{c^2}H'_y\right); \\ &H_x = H'_x, \quad H_y = \Gamma(H'_y - VD'_z), \qquad H_z = \Gamma(H'_z + VD'_y). \end{aligned} \tag{6.37}$$

Exactly the same Eqs. (6.36) and (6.37) are of course obtained in the real 4-space. We shall not mention this space any more because hereinafter we shall only use the final equations and they are identical; moreover, substantial difference is noticed only in transition from Eq. (6.27) to (6.27′). What follows is a simple matter.

It is seen from Eq. (6.36) that all field vectors change their magnitude and direction on transition from one inertial frame of reference K' to another K. Only "longitudinal components" remain invariable, i.e. components along the direction of the relative motion (along the x axis).

Let us expand the electric and magnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$ into the components parallel and perpendicular to the motion direction (the unit vectors $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ are directed along the x, y, z axes respectively), e.g.

$$\boldsymbol{E}_\parallel = E_x\boldsymbol{i}, \quad \boldsymbol{E}_\perp = E_y\boldsymbol{j} + E_z\boldsymbol{k}.$$

Having noted that the velocity vector $\boldsymbol{V}$ of the coordinate system K' has the components $(V, 0, 0)$, we obtain

$$[\boldsymbol{V}\boldsymbol{B}'] = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ V & 0 & 0 \\ B'_x & B'_y & B'_z \end{vmatrix} = -\boldsymbol{j}VB'_z + \boldsymbol{k}VB'_y = V(-\boldsymbol{j}B'_z + \boldsymbol{k}B'_y),$$

$$[\boldsymbol{V}\boldsymbol{E}'] = V(-\boldsymbol{j}E'_z + \boldsymbol{k}E'_y).$$

Then Eq. (6.36) can be rewritten in the vector form:

$$\begin{aligned} \boldsymbol{E}_\parallel &= (\boldsymbol{E}' - [\boldsymbol{V}\boldsymbol{B}'])_\parallel, & \boldsymbol{E}_\perp &= \Gamma(\boldsymbol{E}'_\perp - [\boldsymbol{V}\boldsymbol{B}'])_\perp; \\ \boldsymbol{B}_\parallel &= \left(\boldsymbol{B}' + \tfrac{1}{c^2}[\boldsymbol{V}\boldsymbol{E}']\right)_\parallel, & \boldsymbol{B}_\perp &= \Gamma\left(\boldsymbol{B}'_\perp + \tfrac{1}{c^2}[\boldsymbol{V}\boldsymbol{E}']\right)_\perp. \end{aligned} \tag{6.38}$$

It is appropriate perhaps to recall here that all expressions of the type $[\boldsymbol{V}\boldsymbol{A}]_\parallel$ are equal to zero while expressions of the type $[\boldsymbol{V}\boldsymbol{A}]_\perp$ coincide with the vector product itself for any $\boldsymbol{A}$. The reverse transformation equations are obtained by substitution of unprimed quantities for primed ones and vice versa, and by changing the sign of V to the opposite:

$$\begin{aligned} \boldsymbol{E}'_\parallel &= (\boldsymbol{E} + [\boldsymbol{V}\boldsymbol{B}])_\parallel, & \boldsymbol{E}'_\perp &= \Gamma(\boldsymbol{E} + [\boldsymbol{V}\boldsymbol{B}])_\perp; \\ \boldsymbol{B}'_\parallel &= \left(\boldsymbol{B} - \tfrac{1}{c^2}[\boldsymbol{V}\boldsymbol{E}]\right)_\parallel, & \boldsymbol{B}'_\perp &= \Gamma\left(\boldsymbol{B} - \tfrac{1}{c^2}[\boldsymbol{V}\boldsymbol{E}]\right)_\perp. \end{aligned} \tag{6.39}$$

In the case of non-relativistic velocities $\Gamma \approx 1$, and we obtain from Eq. (6.38)

$$\boldsymbol{E} = \boldsymbol{E}' + [\boldsymbol{B}'\boldsymbol{V}], \quad \boldsymbol{B} = \boldsymbol{B}' - \tfrac{1}{c^2}[\boldsymbol{E}'\boldsymbol{V}]. \tag{6.40}$$

The following designations are used: $\boldsymbol{E} = \boldsymbol{E}_\parallel + \boldsymbol{E}_\perp$ and $\boldsymbol{B} = \boldsymbol{B}_\parallel + \boldsymbol{B}_\perp$. The equations of reverse transformation from K to K' are obtained as usual by substitution of unprimed quantities for primed ones and vice versa with a simultaneous change of sign of V:

$$\boldsymbol{E}' = \boldsymbol{E} + [\boldsymbol{V}\boldsymbol{B}], \quad \boldsymbol{B}' = \boldsymbol{B} - \tfrac{1}{c^2}[\boldsymbol{V}\boldsymbol{E}]. \tag{6.41}$$

In conclusion let us write out the transformation equations for $\boldsymbol{D}$ and $\boldsymbol{H}$. One may not compute anything; it is sufficient to recall that we have obtained the transformation equations for the components of the tensor $\mathfrak{F} = (c\boldsymbol{B}, -i\boldsymbol{E})$, and now we are interested in analogous equations for the tensor $f = (\boldsymbol{H}, -ic\boldsymbol{D})$. Instead of

Eq. (6.39) we obtain the following expressions for the corresponding components·

$$\begin{aligned} D'_{\parallel} &= \left(D + \frac{1}{c^2}[VH]\right)_{\parallel}, & D'_{\perp} &= \Gamma\left(D + \frac{1}{c^2}[VH]\right)_{\perp}; \\ H'_{\parallel} &= (H - [VD])_{\parallel}, & H'_{\perp} &= \Gamma(H - [VD])_{\perp}. \end{aligned} \tag{6.42}$$

In the case of non-relativistic velocities, when $\Gamma \approx 1$, Eq. (6.41) turns into

$$D' = D + \frac{1}{c^2}[VH], \quad H' = H - [VD]. \tag{6.43}$$

Suppose that in the frame K' the magnetic field $B' = 0$. Then in the frame K the relationship between E and B becomes very simple. First of all, notice that $[VE] = [VE_{\perp}]$ since $[VE_{\parallel}] = 0$. From Eq. (6.38) we obtain

$$\begin{aligned} &E = E'_{\parallel} + \Gamma E'_{\perp}, \\ B = \Gamma\frac{1}{c^2}[VE'] = \frac{\Gamma}{c^2}[VE'_{\perp}] = \frac{1}{c^2}[V, \Gamma E'_{\perp}] &= \\ &= \frac{1}{c^2}[V, E'_{\parallel} + \Gamma E'_{\perp}] = \frac{1}{c^2}[VE]. \end{aligned} \tag{6.44}$$

Similarly, if in the frame K' the field E' is equal to zero or in the frame K the field E is equal to zero, then

$$E = -[VB], \quad E' = [VB']. \tag{6.45}$$

In both cases and in any inertial frame the fields turn out to be mutually perpendicular. It follows from both the relativistic equations (6.38) and the approximate equations (6.41) for low velocities that if in one of the frames (say, K) an electric or a magnetic field is equal to zero, electric and magnetic fields in all other inertial frames of reference are perpendicular to each other. The same result can be obtained by employing the Lorentz transformation invariants (see § 6.5).

If the fields E' and B' are mutually perpendicular in a reference frame K', there exists a reference frame K in which one of the fields disappears. It will be shown in § 6.5 that the expression $c^2B^2 - E^2$ remains invariant under the Lorentz transformation. Consequently, if the condition $c^2B'^2 - E'^2 < 0$ is satisfied in the frame K', one can obtain a purely electric field through the appropriate choice of a reference frame, while in the case of $c^2B'^2 - E'^2 > 0$ one gets a purely magnetic field. We shall show how to find the velocity V of the reference frame K. Suppose this velocity is perpendicular to B' in the case of $c^2B'^2 - E'^2 < 0$ and to E' in the case of $c^2B'^2 - E'^2 > 0$. Then $B_{\parallel} = 0$ in the former case and $E_{\parallel} = 0$ in the latter case. Now one has to ensure that

$\boldsymbol{B}_\perp = 0$ in the former case; in order to do this, the following condition should be met (see Eq. (6.38)):

$$\boldsymbol{B}'_\perp + (1/c^2)[\boldsymbol{V}\boldsymbol{E}']_\perp = 0.$$

Multiplying both sides of this expression by $\boldsymbol{E}'$ vectorwise and taking into account the relations $[\boldsymbol{V}\boldsymbol{E}']_\perp = [\boldsymbol{V}\boldsymbol{E}'_\perp]_\perp = [\boldsymbol{V}\boldsymbol{E}'_\perp]$, $[\boldsymbol{E}'[\boldsymbol{V}\boldsymbol{E}'_\perp]]_\perp = \boldsymbol{V}E'^2$, $\boldsymbol{B}'_\perp = \boldsymbol{B}'$, we obtain the reference frame velocity $\boldsymbol{V}$

$$\boldsymbol{V} = -(c^2/E'^2)[\boldsymbol{E}'\boldsymbol{B}']. \tag{6.46}$$

In much the same manner, we obtain in the other case

$$\boldsymbol{V} = (1/B'^2)[\boldsymbol{E}'\boldsymbol{B}']. \tag{6.47}$$

One can always find such an inertial frame of reference in which an electric and a magnetic field are parallel to each other at a given point (see, however, the comment on light waves at the end of § 6.5). Obviously, provided there exists one such frame, there should be an infinite number of frames possessing the same property. In fact, in any inertial frame of reference K' moving rectilinearly and uniformly relative to K in the direction coinciding with the common direction of $\boldsymbol{E}$ and $\boldsymbol{B}$, the fields $\boldsymbol{E}'$ and $\boldsymbol{B}'$ will remain parallel since the field components oriented along the motion direction do not vary.

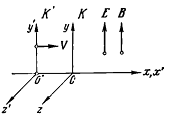


Fig. 6.2. A transition to the reference frame K in which an electric and a magnetic field turn out to be parallel.

In order to find at least one frame in which the fields are parallel, we shall proceed as follows. Suppose that the fields are parallel in the frame K, i.e. $[\boldsymbol{E}\boldsymbol{B}] = 0$. Direct the velocity of the frame K' (in which the fields $\boldsymbol{E}'$ and $\boldsymbol{B}'$ are not parallel any more) along the perpendicular to the fields $\boldsymbol{E}$ and $\boldsymbol{B}$; assume that the x, x' axis is directed along the velocity $\boldsymbol{V}$ (see Fig. 6.2). Then $E_x = B_x = 0$ and, since the vector cross product is equal to zero, $E_yB_z - E_zB_y = 0$. Substituting into this equation the components of $\boldsymbol{E}$ and $\boldsymbol{B}$, expressed via the components of $\boldsymbol{E}'$ and $\boldsymbol{B}'$ according to Eq. (6.36), we arrive at the following equation:

$$\Gamma(E'_y + VB'_z)\Gamma\left(B'_z + \frac{V}{c^2}E'_y\right) = \Gamma(E'_z - VB'_y)\Gamma\left(B'_y - \frac{V}{c^2}E'_z\right).$$

The frame velocity V can be determined from this equation using the given fields $\boldsymbol{E}'$ and $\boldsymbol{B}'$. Taking into account that according to Eq. (6.36) $E'_x = B'_x = 0$, we can immediately find the direction

of the velocity V relative to $\boldsymbol{E}'$ and $\boldsymbol{B}'$. Indeed, $[\boldsymbol{E}'\boldsymbol{B}'] = \boldsymbol{i}(E'_yB'_z - - E'_zB'_y)$ and $\boldsymbol{V} = V\cdot\boldsymbol{i}$, so that solving the foregoing equation, one can write

$$\frac{V/c^2}{1+V^2/c^2} = -\frac{[\boldsymbol{E}'\boldsymbol{B}']}{c^2B'^2+E'^2}. \tag{6.48}$$

Thus, from the given vectors $\boldsymbol{E}'$ and $\boldsymbol{B}'$ in the frame K' one can find the frame K in which $\boldsymbol{E}$ and $\boldsymbol{B}$ are parallel. The velocity direction of this frame coincides with that of $[\boldsymbol{E}'\boldsymbol{B}']$, while the velocity magnitude is one of the roots of the quadratic equation (6.48). Surely, from the two roots of Eq. (6.48) one should choose the one for which $V < c$. The case $\boldsymbol{E}'\boldsymbol{B}' = 0$ was examined above. one cannot obtain parallel fields here, but it is possible to get either a purely magnetic or a purely electric field.

§ 6.5. The electromagnetic field invariants. Although an electric field strength $\boldsymbol{E}$ and a magnetic field induction $\boldsymbol{B}$ vary under the Lorentz transformation, there are some combinations of these fields remaining invariable under it. These quantities are invariants of antisymmetric 4-tensors of the second rank. We make use of two such invariants (see Appendix I, § 6):

$$I_1 = F_{ik}^2, \quad I_2 = F_{ik}F^*_{ik} = \frac{1}{2}e_{iklm}F_{ik}F_{lm}.$$

Recalling the definitions of the tensors F_{ik} and F^*_{ik}

$$\mathfrak{F}(c\boldsymbol{B}, \; -i\boldsymbol{E}), \quad \mathfrak{F}^*(-i\boldsymbol{E}, \; c\boldsymbol{B})$$

and taking into account that the first invariant is the sum of the components F_{ik} squared and the second invariant represents the pairwise products of the corresponding components of the tensors F_{ik} and F^*_{ik}, we can write at once $I_1 = 2(c^2B^2 - E^2)$, $I_2 = = -2ic(\boldsymbol{BE})$.

Omitting immaterial constant factors, one can claim that an electromagnetic field possesses two invariants (we shall not write out the invariants of the tensor $\mathfrak{F}$ and the combined invariants of $\mathfrak{F}$ and $\mathfrak{f}$ since they will not be needed):

$$I_1 = c^2B^2 - E^2, \quad I_2 = \boldsymbol{BE}.$$

From the existence of these two invariants follow the results, some of which have been mentioned before. If in some IFR the fields $\boldsymbol{E}$ and $\boldsymbol{B}$ are mutually orthogonal ($[\boldsymbol{EB}] = 0$), they are also orthogonal in any other inertial frame of reference. If in some reference frame $E = cB$, this relationship holds in all inertial frames of reference.

It should be noted here that both invariants are equal to zero for a light wave *in vacuo*. These properties, i.e. $\boldsymbol{B} \perp \boldsymbol{E}$ and $cB = E$, are maintained in any IFR.

It is clear that if $I_2 = 0$ and $I_1 \neq 0$, one can always find a reference frame in which either $\boldsymbol{E} = 0$ or $\boldsymbol{B} = 0$ (depending on the sign of I_1), i.e. pass over to either a purely magnetic or a purely electric field. Conversely, if either $\boldsymbol{E}$ or $\boldsymbol{B}$ is equal to zero in some frame, these fields will be mutually orthogonal in all other inertial frames. Note that the quantity $\boldsymbol{BE}$ is not a "real" scalar since it changes sign on transition from the left coordinate system to the right one and vice versa, while the quantity $(\boldsymbol{BE})^2$ is a real scalar.

§ 6.6. The Lorentz force. Now let us consider the forces acting on electric charges in an electromagnetic field. To avoid confusion, we shall confine our presentation to spatial distribution of charges *. In a co-moving reference frame K' in which a considered space element rests together with a charge, this charge experiences a force exerted by an electric field (a magnetic field does not act on a charge at rest). The force acting on a charge contained in a unit volume is referred to as a *force density*. If a charge density in a co-moving reference frame K' is equal to ρ_0, the force density $\boldsymbol{f}'$ is defined by the equation

$$\boldsymbol{f}' = \rho_0 \boldsymbol{E}',$$

where $\boldsymbol{E}'$ is an electric field strength in K'.

The transition to any other IFR is associated with the variation of the fields $\boldsymbol{E}$ and $\boldsymbol{B}$; even if in a co-moving frame there was no magnetic field and only an electric one was present, a magnetic field will appear in any other IFR. Let us find the force density $\boldsymbol{f}'$ expressed via the field components $\boldsymbol{E}$ and $\boldsymbol{B}$ in an arbitrary inertial frame. First, let us examine the case of non-relativistic velocities when $\Gamma \approx 1$; then according to Eq. (6.17) $\rho = \rho_0 \Gamma \approx \rho_0$, and according to Eq. (6.41) $\boldsymbol{E}' = \boldsymbol{E} + [\boldsymbol{VB}]$, and because of this

$$\boldsymbol{f} = \boldsymbol{f}' = \rho_0 \boldsymbol{E}' = \rho \{\boldsymbol{E} + [\boldsymbol{VB}]\}. \tag{6.49}$$

The last link of Eq. (6.49) defines the quantity which is usually called the *Lorentz force density* in electrodynamics. The Lorentz force defines the force acting on a unit volume containing a charge; this force is generated by the electric and magnetic fields of the frame K relative to which the charge moves at the velocity $\boldsymbol{V}$. It is not surprising that the force $\boldsymbol{f}'$ in the frame K' turned out to be equal to the force $\boldsymbol{f}$ in the frame K since according to Eq. (5.34) a force magnitude does not change on transition from one IFR to another in a non-relativistic case.

Of course, Eq. (6.49) can also be used in the case when the velocity of charge motion is different at various points in space. In this case each element of space will have its own co-moving

* Point charges are discussed, for example, in [8], § 29.

reference frame and, consequently, the velocity V will be different at various points.

Let us derive the expression for the Lorentz force by still another method illustrating explicitly how Eq. (6.49) comes about. Let the co-moving frame K' have an electric and a magnetic field defined by the vectors E' and B'.

Making use of the superposition principle, we can describe each of these fields as a sum of two fields:

$$\text{I}\quad E'_1 = 0, \qquad B'_1 = B';$$
$$\text{II}\quad E'_2 = E', \qquad B'_2 = 0.$$

Obviously, the initial field represents just the sum of the two fields: $E' = E'_1 + E'_2$, $B' = B'_1 + B'_2$. However, the field transformation equations I and II are very simple and allow us to get the answer right away. In the frame K'

$$f' = \rho_0 E' = \rho_0 E'_2. \tag{6.50}$$

Using the first equation of (6.45), one can write down immediately the electric field I in the frame K:

$$E_1 = -[VB_1],$$

where B_1 is the magnetic field in K. According to Eq. (6.36) the electric field II in the frame K is equal to

$$E_2 = E'_{2x}i + \Gamma(E'_{2y}j + E'_{2z}k) \sim E'_2$$

in the case when $\Gamma \approx 1$. The total electric field in K is equal to the sum of E_1 and E_2:

$$E = E_1 + E_2 = E'_2 - [VB_1]. \tag{6.51}$$

The magnetic field B in the frame K is equal to $B_1 + B_2$. Having composed the vector cross product $[VB] = [VB_1] + [VB_2]$ we see from the second equation of (6.44) defining B_2 that the product $|[VB_2]| \approx (V^2/c^2)$ can be ignored in a non-relativistic case. Therefore, $[VB_1] = [VB]$ and we obtain the Lorentz force (see Eq. (6.49)) from Eq. (6.51).

If in the frame K an electric field is equal to zero ($E = 0$) and a magnetic field differs from zero, it follows from Eq. (6.45) that $E' = [VB']$; so, the Lorentz force, appearing to be produced by a pure magnetic field in the frame K, seems to be produced in a co-moving frame K' by a pure electric field. These examples show once again a uniqueness of an electromagnetic field and the relativity of its division into an electric and a magnetic field.

A few words on lines of force of the field are relevant here. In each reference frame a vector field can be correlated with a family of vector lines of force. These lines are formally defined as curves

whose tangents at every point coincide with the direction of the field vector at that point. A line of force is a useful assisting notion allowing the properties of the field to be graphically exhibited. In contrast to the ideas of the last century, however, no one attaches any physical meaning to these lines now.

Suppose, a charge or a constant magnet moves in space. Should one say in this case that a field and its lines of force move along?

A field is a method to describe what happens at a given point in space. The magnet motion merely causes the field to vary in

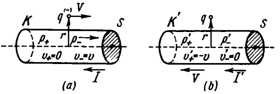


Fig. 6.3. The interaction of a charge q, moving at the velocity V parallel to a current-carrying conductor, and a current. (*a*) A conductor is at rest in the frame K, the charge and electrons move at the velocity V. (*b*) A conductor moves at the velocity V in the frame K', the charge and electrons are at rest.

time at a given point. And still, one may speak of the field motion induced by a charge or magnet moving at a constant velocity, since this field moves with them as a whole. The field transportation velocity is the velocity at which a charge or a magnet moves. A motion of lines of force, however, is better not to be mentioned since a motion velocity of lines of force has no physical meaning. An auxiliary nature of lines of force is demonstrated particularly well by the fact that they may just disappear in some reference frame for a certain field.

Here is another example illustrating the relative character of forces acting in an electromagnetic field. Consider a cylindrical conductor carrying a current and a negative charge q moving parallel to the conductor at the velocity V (Fig. 6.3). We shall fix the frame K to the conductor, and the frame K' to the charge. In the frame K the charge experiences the Lorentz force induced by a magnetic field and directed at right angles to the conductor's axis. Consequently, the charge approaches the conductor. In the frame K', however, the charged particle is at rest and the magnetic field has no effect on it. Then, what is the reason causing the charge to deviate in terms of the frame K'?

Here one needs to review a microscopic description of what is happening in a conductor. A current originates in a conductor due to the motion of free electrons since positive ions and fixed (valence) electrons cannot migrate along a conductor. Let the

density of conduction electrons be equal to ρ_-, with their velocity in K (relative to the conductor) equal to v_-. The density of stationary conduction electrons is equal to ρ_+, and due to the neutrality of the conductor $\rho_+ + \rho_- = 0$. Inasmuch as the conductor is neutral, there is no electric field outside of it, and the force acting on the charge q arises only from a magnetic field:

$$\boldsymbol{F} = q\,[\boldsymbol{V}\boldsymbol{B}]:$$

The magnitude of the magnetic field induced by a rectilinear current at the distance r from its axis is known:

$$B = \frac{\mu_0 I}{2\pi r},$$

the vector $\boldsymbol{B}$ coincides with the tangent of a circle lying in a plane perpendicular to the current's axis and having its centre on it. The direction of the vector $\boldsymbol{B}$ is determined by the right-hand screw rule. Hence, the force acting on the charge is directed toward the conductor and is equal to

$$F = \frac{qV\mu_0 I}{2\pi r} = \frac{1}{4\pi\varepsilon_0 c^2}\,\frac{2IqV}{r}.$$

The current can be expressed via the conduction electron velocity v_-, their density and the cross-sectional area S:

$$I = jS = \rho_- v_- S,$$

whence

$$F = \frac{q}{2\pi\varepsilon_0}\,\frac{\rho_- S}{r}\,\frac{Vv}{c^2} = \frac{q}{2\pi\varepsilon_0}\,\frac{\rho_+ S}{r}\,\frac{V^2}{c^2}, \tag{6.52}$$

provided that for the sake of simplicity we assume the velocity of electrons in metal to be equal to that of the charge q, i.e. $V = v_-$.

Now let us consider the same situation in the frame K'. The charge q and conduction electrons are at rest in K'. This time, however, the charges connected with the conductor (and whose density is equal to ρ_+) move relative to the charge q. Although they induce a certain magnetic field B', it does not act on the charge q any more, since the charge is motionless in K'. Whence it is clear at once that an electric field must appear in the frame K' since the charge must deviate toward the axis in K' as well. Its origin is easy to understand from the results obtained by us earlier. Conduction electrons are at rest in the frame K', and therefore $\rho_- = \Gamma\rho'_-$ (see Eq. (6.17)). Positive charges connected with the conductor move at the velocity $-V$ in the frame K', and so $\rho'_+ = \Gamma\rho_+$ (these charges were stationary in K). The resulting

charge density ρ' is equal to $\rho'_+ + \rho'_-$ in the frame K', and, consequently,

$$\rho' = \rho_-/\Gamma + \Gamma\rho_+ = \rho_+(\Gamma - 1/\Gamma) = \Gamma\rho_+ B^2,$$

where the relation $\rho_+ = -\rho_-$ is allowed for; this equation coincides with Eq. (6.24). Consequently, a moving conductor is charged positively with the space density ρ'. But an electric field

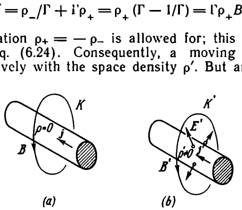


Fig. 6.4. (*a*) In the frame K the charge density ρ ($\rho = \rho_+ + \rho_-$) is equal to zero while the current density is equal to $j \neq 0$. Therefore, an electric field is absent in the frame K and there is only a magnetic field $\boldsymbol{B}$. (*b*) In the frame K' the density of charges that emerges is equal to ρ', and the current density becomes equal to j'. A magnetic field is equal to $\boldsymbol{B}'$; besides, an electric field $\boldsymbol{E}'$ also appears.

of a uniformly charged cylinder is also known from electrodynamics. It falls within the planes perpendicular to the cylinder's axis and is oriented along the rays leaving the cylinder's axis. Its magnitude

$$E' = \frac{\rho' S}{2\pi\varepsilon_0 r} = \frac{\rho_+ S\Gamma B^2}{2\pi\varepsilon_0 r}.$$

This implies that the force acting on a negative charge q is directed toward the conductor, and its magnitude in the frame K' is equal to

$$F' = qE' = \frac{q}{2\pi\varepsilon_0}\frac{\rho_+ S}{r}\Gamma B^2.$$

Comparing this result with Eq. (6.52), we see that these forces are equal in a non-relativistic approximation ($\Gamma \approx 1$). Recalling that the forces transform according to Eq. (5.34), we find that both ways of describing an observed phenomenon give identical results at any velocity V. The results pertaining to fields in the frames K and K' are explained in Fig. 6.4.

In conclusion it should be emphasized that all the results pertaining to forces which an electromagnetic field exerts on space charges, are obtained quite easily provided that the Lorentz force

density (see Eq. (6.49))

$$\mathbf{f} = \rho\{\mathbf{E} + [\mathbf{vB}]\}$$

is written in a four-dimensional form. To pass over to the four-dimensional notation, rewrite the x-component of the Lorentz force as follows:

$$f_x = f_1 = \rho E_x + \rho v_y B_z - \rho v_z B_y =$$
$$= \left(-\frac{is_4}{c}\right) iF_{14} + s_2 \frac{F_{12}}{c} - s_3\left(-\frac{F_{13}}{c}\right) =$$
$$= \frac{1}{c}(F_{12}s_2 + F_{13}s_3 + F_{14}s_4) = \frac{1}{c} F_{1k}s_k.$$

The following relations are taken into account in this chain of equations:

$$\rho = -\frac{i}{c}s_4, \quad \rho v_y = s_2, \quad \rho v_z = s_3,$$
$$B_z = \frac{F_{12}}{c}, \quad B_y = -\frac{F_{13}}{c}, \quad F_{11} = 0.$$

Analogous expressions are obtained for $f_y = f_2$ and $f_z = f_3$. Hence it is clear that the 4-vector of a force density acting on a charge in an electromagnetic field, to be denoted by $\vec{f}$, has the components *

$$f_i = \frac{1}{c} F_{ik}s_k. \tag{6.53}$$

We have already pointed out that the separation of forces exerted on the charge by electric and magnetic fields into the parts $\rho\mathbf{E}$ and $\rho[\mathbf{vB}]$ is relative. Both these forces constitute a united whole combining naturally into a single four-dimensional expression (Eq. (6.53)).

As we have seen, the first three density components lead to the conventional three-dimensional equation (6.49). Let us find the fourth component:

$$f_4 = \frac{1}{c} F_{4k}s_k = \frac{1}{c}(F_{41}s_1 + F_{42}s_2 + F_{43}s_3) = \frac{i\rho}{c}(\mathbf{vE}).$$

The quantity $\rho(\mathbf{vE})$ has a simple meaning which is immediately seen as soon as the two sides of Eq. (6.49) are multiplied scalarwise by $\mathbf{v}$. Taking into consideration that $[\mathbf{vB}]\mathbf{v} = 0$, we get

$$(\mathbf{fv}) = \rho(\mathbf{vE}).$$

* We expect that the reader will not forget that the letter F supplemented with two subindices represents a tensor $\mathfrak{f}$ component. The components of the 4-force density have a single subindex.

The left-hand side of the last equation represents the power of a Lorentz force per unit of volume (forces exerted by a magnetic field perform no work):

$$f_4 = \frac{i}{c}(\boldsymbol{f}\boldsymbol{v}).$$

Thus we have obtained a 4-force density vector whose components are written down together as follows:

$$\vec{f}\begin{Bmatrix} f_1 & f_2 & f_3 & f_4 \\ f_x & f_y & f_z & \frac{i}{c}(\boldsymbol{f}\boldsymbol{v}) \end{Bmatrix}. \qquad (6.54)$$

Let us consider the force exerted by an electromagnetic field on a unit of volume containing a charge ρ_0 in the reference frame K^0 co-moving with the charge. Then $\vec{f}(\rho_0\boldsymbol{E}', 0)$. Passing over to any other reference frame K, we get

$$f_1 = f_1', \quad f_2 = \Gamma f_2', \quad f_3 = \Gamma f_3', \quad f_4 = -i\mathrm{B}\Gamma f_1' = -i\frac{V}{c}\Gamma\rho_0 E_x'.$$

Here the equations for a force density in the frame K are expressed through the fields in the frame K'. Usually a force density is expressed in terms of the quantities referred to the frame in which the force density is determined. Making use of Eqs. (6.17) and (6.36), we obtain $\vec{f}(\rho(\boldsymbol{E}+[\boldsymbol{V}\boldsymbol{B}]),\ i/c\ \rho(\boldsymbol{E}\boldsymbol{v}))$ for non-relativistic velocities (ignoring the terms V^2/c^2).

In conclusion we shall write out the motion equation for a charged particle in a four-dimensional form:

$$\frac{d}{d\tau}(mu_i) = \frac{1}{c}F_{ik}s_k. \qquad (6.55)$$

§ 6.7. Covariance of the system of the Maxwell equations. The Maxwell equations define the behaviour of an electromagnetic field in the most adequate manner. They were proposed long before the advent of the theory of relativity and surely before the Lorentz transformation was identified. According to the principle of relativity the appearance of the Maxwell equations must remain constant in all inertial frames of reference. Consequently, the Maxwell equations must be covariant relative to the Lorentz transformation. It so happened that the system of Maxwell's equations satisfies these conditions when written in the form proposed by its creator. To ascertain this, the system of Maxwell's equations, written usually in terms of three-dimensional equations, should be rewritten in a four-dimensional form. Now we shall be occupied with just that. The system of the Maxwell equations is known

to have the following form:

$$\operatorname{rot} \boldsymbol{H} = j + \dot{\boldsymbol{D}}, \quad \text{(a)} \quad \Big| \quad \operatorname{div} \boldsymbol{D} = \rho; \quad \text{(b)} \tag{6.56}$$

$$\operatorname{rot} \boldsymbol{E} = -\dot{\boldsymbol{B}}, \quad \text{(a)} \quad \Big| \quad \operatorname{div} \boldsymbol{B} = 0. \quad \text{(b)} \tag{6.57}$$

We have split the equations into two lines, having combined the equations involving the average values of an electric and a magnetic field $\boldsymbol{E}$ and $\boldsymbol{B}$, and the equations for the subsidiary vectors $\boldsymbol{H}$ and $\boldsymbol{D}$.

In order to present Eqs. (6.56) and (6.57) in a four-dimensional form, we shall need the tensors (6.29a) and (6.31); we shall also make use of the definition of a 4-current density vector (6.12a). Note for the future use, by the way, that tensors $\mathfrak{F}$ and $\mathfrak{f}$ are linked *in vacuo* by the relation

$$\mathfrak{F} = \sqrt{\frac{\mu_0}{\varepsilon_0}}\, \mathfrak{f}. \tag{6.58}$$

Naturally, Eqs. (6.56) can be expressed via tensor (6.31) while Eqs. (6.57) via tensor (6.29a).

Let us consider the x component of Eq. (6.56a):

$$\dot{D}_x - \frac{\partial H_z}{\partial y} + \frac{\partial H_y}{\partial z} = -j_x. \tag{6.59}$$

Recalling that according to Eq. (6.12a) $j_x = s_1$ and using the first line of Eq. (6.31) together with the definitions $x_1 = x$, $x_2 = y$, $x_3 = z$, $x_4 = ict$, we shall rewrite Eq. (6.59) as $-\frac{\partial f_{12}}{\partial x_2} - \frac{\partial f_{13}}{\partial x_3} - \frac{\partial f_{14}}{\partial x_4} = -s_1$. The two other components are given by similar expressions which can be written in a general form as ($i = 1, 2, 3$):

$$\frac{\partial f_{ik}}{\partial x_k} = s_i, \tag{6.60}$$

where the summation is performed over k running from 1 to 4. It is readily seen that when $i = 4$, we get Eq. (6.56b). Thus Eq. (6.56) is rewritten in the form of Eq. (6.60), but now in terms of 4-tensor components (6.31).

Now let us consider the components of Eq. (6.57). For example, the x component of Eq. (6.57a) can be written as

$$\dot{B}_x + \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = 0. \tag{6.61}$$

Resorting to tensor (6.29a), one can rewrite Eq. (6.61) as follows:

$$i\left(\frac{\partial F_{23}}{\partial x_4} + \frac{\partial F_{34}}{\partial x_2} + \frac{\partial F_{42}}{\partial x_3}\right) = 0. \tag{6.62}$$

It is not difficult to notice that consecutive terms in Eq. (6.62) are obtained via a cyclic transposition of the three indices in each of the preceding terms. The structure of Eq. (6.62), however, becomes quite evident if we introduce the tensor $F^{\bullet}_{ik}$ which is dual to the tensor F_{ik} (see Appendix I, § 6):

$$F^{\bullet}_{ik} = \frac{1}{2} e_{iklm} F_{lm}, \tag{6.63}$$

where e_{iklm} is a fully antisymmetric unit 4-tensor of the fourth rank. One can easily see that the dual tensor $F^{\bullet}_{ik}$ differs from the tensor F_{ik} only by transposed components of the imaginary and real parts:

$$\mathfrak{F} = (c\boldsymbol{B},\ -i\boldsymbol{E}),\quad \mathfrak{F}^{\bullet} = (-i\boldsymbol{E},\ c\boldsymbol{B}), \tag{6.64}$$

or, written in full,

$$F^{\bullet}_{ik} = \begin{pmatrix} 0 & -iE_z & iE_y & cB_x \\ iE_z & 0 & -iE_x & cB_y \\ -iE_y & iE_x & 0 & cB_z \\ -cB_x & -cB_y & -cB_z & 0 \end{pmatrix}. \tag{6.65}$$

Using this tensor, the pair of Maxwell's equations (6.57) can be rewritten in a four-dimensional form as follows:

$$\frac{\partial F^{\bullet}_{ik}}{\partial x_k} = 0. \tag{6.66}$$

Let us make sure that Eq. (6.66) corresponds to the four equations of (6.57).

Eq. (6.66) contains four equations ($i = 1, 2, 3, 4$). Consider, for example, the equation for $i = 1$:

$$\frac{\partial F^{\bullet}_{1k}}{\partial x_k} = \frac{\partial F^{\bullet}_{11}}{\partial x_1} + \frac{\partial F^{\bullet}_{12}}{\partial x_2} + \frac{\partial F^{\bullet}_{13}}{\partial x_3} + \frac{\partial F^{\bullet}_{14}}{\partial x_4} =$$

$$= -i\frac{\partial E_z}{\partial y} + i\frac{\partial E_y}{\partial z} + \frac{\partial (cB_x)}{\partial (ict)} = 0,$$

or otherwise

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\frac{\partial B_x}{\partial t}, \quad \text{i. e.} \quad (\operatorname{rot} \boldsymbol{E})_x = -(\dot{\boldsymbol{B}})_x.$$

Eq. (6.66) yields the two other components of the equation $\operatorname{rot} \boldsymbol{E} = -\dot{\boldsymbol{B}}$ at $i = 2, 3$. For $i = 4$ we get the following equation:

$$\frac{\partial F^{\bullet}_{4k}}{\partial x_k} = \frac{\partial F^{\bullet}_{41}}{\partial x_1} + \frac{\partial F^{\bullet}_{42}}{\partial x_2} + \frac{\partial F^{\bullet}_{43}}{\partial x_3} - \frac{\partial F^{\bullet}_{44}}{\partial x_4} =$$

$$= -\frac{\partial (cB_x)}{\partial x} - \frac{\partial (cB_y)}{\partial y} - \frac{\partial (cB_z)}{\partial z} = 0;$$

representing Eq. (6.57b): $\operatorname{div} \boldsymbol{B} = 0$. So we see that Eq. (6.66) comprises Maxwell's equations (6.57).

Quite often Eq. (6.57) is expressed directly through the tensor F_{ik}. We present this notation here since it will be necessary later on. The Maxwell equations (6.57) can be legitimately written both in the form of Eq. (6.66) and as an equation of the type obtained in (6.62):

$$\frac{\partial F_{ik}}{\partial x_l} + \frac{\partial F_{kl}}{\partial x_i} + \frac{\partial F_{li}}{\partial x_k} = 0. \tag{6.67}$$

There is no summation involved in Eq. (6.67). Three different values of the indices i, k, l are to be chosen from the four possible ones. The reader can make sure himself that if two of these indices are taken to be the same, and antisymmetric properties of the tensor $\mathfrak{F}(F_{ik} = -F_{ki})$ are allowed for, Eq. (6.67) is seen to turn into an identity. The structure of Eq. (6.67) shows that the distribution of the chosen triad of numbers among the indices i, k, l is insignificant. This implies that Eq. (6.67) contains several independent equations, their number being equal to the number of possible combinations, each containing three indices, that can be formed from a collection of four indices, that is ${}_4C_3 = {}_4C_1 = 4$. We let the reader make sure for himself that the four equations (6.57) follow from Eq. (6.67).

Now the Maxwell equations can be readily proved to be covariant. We have seen that they can be written as Eq. (6.60) and (6.66) or (6.60) and (6.67). But Eqs. (6.60) and (6.66) represent the relations between 4-vectors since the expression $\partial f_{ik}/\partial x_k$ is a vector (see Appendix I, § 5). Eq. (6.60) differs from Eq. (6.66) by the zero vector featured on the right-hand side of Eq. (6.66). As to Eq. (6.67), it is explicitly presented in a tensor form and consequently is covariant. Thus, the four-dimensional notation of Maxwell's equations itself indicates their covariance.

The system of Maxwell's equations is not confined to Eqs. (6.56) and (6.57). We have already quoted the charge conservation law in a covariant form (see p. 184). It remains only to rewrite the "material equations" in a covariant form.

§ 6.8. The Minkowski equations for moving media (the transformation of material equations). In the previous section we saw that the system of Maxwell's equations (6.56) and (6.57) retains its appearance in all inertial frames of reference. However, the Maxwell equations yield an unambiguous picture of electromagnetic phenomena only when material equations, characterizing a medium in which electromagnetic phenomena occur, are specified. As usual, the reference frame in which the medium (or its portion) rests will be referred to as a co-moving one. In the case of a uniform isotropic medium the material equations have the following

form in a co-moving frame:

$$D' = \varepsilon E', \tag{6.68}$$

$$B' = \mu H', \tag{6.69}$$

$$j' = \sigma E', \tag{6.70}$$

with a permittivity ε, permeability μ and conductivity σ all being constants. Consider the motion of a medium relative to a "laboratory" frame. In the frame co-moving with a medium the Maxwell equations for a stationary medium are valid. Due to the principle of relativity the material constants ε, μ, σ must be the same both in a stationary medium in a "laboratory" frame and in the reference frame co-moving with the medium. Inasmuch as the transformation equations for the vectors $\boldsymbol{E}$, $\boldsymbol{B}$, $\boldsymbol{H}$ and $\boldsymbol{D}$ are known, the relationship between them can be found in any other inertial frame differing from K'. Let us write out the necessary transformation equations, having split them into the longitudinal and transverse (relative to the reference frame velocity $\boldsymbol{V}$) parts (see § 6.4):

$$E'_{\parallel} = (\boldsymbol{E} + [\boldsymbol{VB}])_{\parallel}, \qquad E'_{\perp} = \Gamma(\boldsymbol{E} + [\boldsymbol{VB}])_{\perp}; \tag{6.71}$$

$$B'_{\parallel} = \left(\boldsymbol{B} - \frac{1}{c^2}[\boldsymbol{VE}]\right)_{\parallel}, \qquad B'_{\perp} = \Gamma\left(\boldsymbol{B} - \frac{1}{c^2}[\boldsymbol{VE}]\right)_{\perp} \tag{6.71'}$$

$$D'_{\parallel} = \left(\boldsymbol{D} + \frac{1}{c^2}[\boldsymbol{VH}]\right)_{\parallel}, \qquad D'_{\perp} = \Gamma\left(\boldsymbol{D} + \frac{1}{c^2}[\boldsymbol{VH}]\right)_{\perp}; \tag{6.72}$$

$$H'_{\parallel} = (\boldsymbol{H} - [\boldsymbol{VD}])_{\parallel}, \qquad H'_{\perp} = \Gamma(\boldsymbol{H} - [\boldsymbol{VD}])_{\perp}. \tag{6.73}$$

We should recall once more that all expressions of the type $[\boldsymbol{VA}]_{\parallel}$ are equal to zero for any $\boldsymbol{A}$ since we deal with projections on the velocity direction and a vector cross product is perpendicular to the velocity $\boldsymbol{V}$. If the corresponding expressions are substituted into Eqs. (6.68) and (6.69), we obtain the identical relationships for both the longitudinal and the transverse components (in which the factor Γ cancels out). These relationships can be combined as follows:

$$\boldsymbol{D} + \frac{1}{c^2}[\boldsymbol{VH}] = \varepsilon(\boldsymbol{E} + [\boldsymbol{VB}]), \tag{6.74}$$

$$\boldsymbol{B} - \frac{1}{c^2}[\boldsymbol{VE}] = \mu(\boldsymbol{H} - [\boldsymbol{VD}]). \tag{6.75}$$

Eqs. (6.74) and (6.75) are called the *Minkowski equations*: ε and μ appearing in these equations represent a permittivity and permeability of a resting medium. These equations differ essentially from Eqs. (6.68) and (6.69) in that they involve all of the field vectors simultaneously. Using Eq. (6.75) one can easily eliminate $\boldsymbol{B}$ from Eq. (6.74) and obtain an equation involving only

the three vectors $\boldsymbol{E}$, $\boldsymbol{D}$, $\boldsymbol{H}$, or eliminate $\boldsymbol{D}$ from Eq. (6.75), using Eq. (6.74).

The equations appear simpler when put down separately in terms of longitudinal and transverse components:

$$\boldsymbol{D}_{\parallel} = \varepsilon \boldsymbol{E}_{\parallel}, \quad \boldsymbol{B}_{\parallel} = \mu \boldsymbol{H}_{\parallel}, \tag{6.76}$$

$$\begin{aligned} \left(1 - \frac{\varepsilon\mu}{\varepsilon_0\mu_0}\frac{V^2}{c^2}\right)\boldsymbol{D}_{\perp} &= \varepsilon\left(1 - \frac{V^2}{c^2}\right)\boldsymbol{E}_{\perp} + (\varepsilon\mu - \varepsilon_0\mu_0)[\boldsymbol{VH}], \\ \left(1 - \frac{\varepsilon\mu}{\varepsilon_0\mu_0}\frac{V^2}{c^2}\right)\boldsymbol{B}_{\perp} &= \mu\left(1 - \frac{V^2}{c^2}\right)\boldsymbol{H}_{\perp} + (\varepsilon\mu - \varepsilon_0\mu_0)[\boldsymbol{VE}]. \end{aligned} \tag{6.77}$$

The first equation of (6.77) is obtained from Eq. (6.74) into which the expression for $\boldsymbol{B}$ is substituted from Eq. (6.75) and then only a transverse component of the relation obtained is taken. In much the same manner one obtains the second equation of (6.77).

It is seen from these equations that if the vectors $\boldsymbol{B}$ and $\boldsymbol{H}$, as well as $\boldsymbol{D}$ and $\boldsymbol{E}$, coincide in direction in an isotropic medium in a co-moving frame K', this is not the case in other reference frames.

Of course, when one examines the motion of a medium, the case of non-relativistic velocities proves to be most interesting. Hence, if one ignores the terms V^2/c^2 and n^2V^2/c^2 compared to unity ($n = \sqrt{\varepsilon\mu/\varepsilon_0\mu_0}$ is a medium refraction index, see Chapter 7) in Eq. (6.77), then Eqs. (6.76) and (6.77) take a simpler form:

$$\boldsymbol{D} = \varepsilon\boldsymbol{E} + \frac{1}{c^2}(n^2 - 1)[\boldsymbol{VH}], \quad \boldsymbol{B} = \mu\boldsymbol{H} + \frac{1}{c^2}(n^2 - 1)[\boldsymbol{VE}]. \tag{6.78}$$

This form of material equations written for a moving medium is employed very often. Due to the equivalence of all IFRs *in vacuo* the last equations turn into Eqs. (6.68) and (6.69) in this case.

It is helpful to rewrite the material equations (6.68)-(6.70) in a four-dimensional tensor form. We shall not derive these equations; we shall just write them out to check that we get Eqs. (6.68)-(6.70) in the frame in which the medium is at rest. Let us introduce a four-dimensional velocity $\vec{V}$ ($\Gamma\boldsymbol{V}$, $ic\Gamma$) for a medium. (Γ is written here since the velocity of an object or a medium V is assumed to be equal to that of the reference frame K'). In a co-moving frame K' the 4-velocity $\vec{V}'$ has the components $U_1' = 0$, $U_2' = 0$, $U_3' = 0$, $U_4' = ic$.

The reader can easily verify thal the tensor equations

$$\frac{1}{c} f_{ik}U_k = \varepsilon F_{ik}U_k, \tag{6.79}$$

$$\frac{1}{c}(F_{ik}U_l + F_{kl}U_i + F_{li}U_k) = \mu(f_{ik}U_l + f_{kl}U_i + f_{li}U_k), \tag{6.80}$$

$$s_i = \frac{\sigma}{c} F_{ik}U_k, \tag{6.81}$$

lead to Eqs. (6.68), (6.69) and (6.70) respectively if $\vec{V}'$ components are substituted into them.

In Eqs. (6.79) and (6.80) the summation is carried out over k. The total number of equations amounts to 4, but since at $i = 4$ we get an identity, there are only three equations, in fact. In Eq. (6.80) one has to find the number of possible combinations of i, k and l that can be formed from four values 1, 2, 3, 4 taken three at a time. This number is equal to ${}_4C_3 = 4$, but since the combination 1, 2, 3 yields an identity, we come back again to three equations as it should be. Having derived the correct expressions for material equations in the frame K', we prove that a tensor notation is also correct.

We shall illustrate an application of a tensor notation by the example of Eq. (6.81). Suppose that a current density is observed in a co-moving frame K' (in which the medium is at rest) and the charge density is equal to zero, i.e. s'_i $(j'_1, j'_2, j'_3, 0)$. The velocity of a medium in the frame K' is equal to $\vec{V}'$ $(0, 0, 0, ic)$. Here are the s'_i components:

$$s'_1 = \frac{\sigma}{c} F'_{1k} U'_k = \frac{\sigma}{c} F'_{14} U'_4 = \frac{\sigma}{c} (-iE'_1)(ic) = \sigma E'_1,$$

i.e. $j'_1 = \sigma E'_1$; similarly $j'_2 = \sigma E'_2$, $j'_3 = \sigma E'_3$. The fourth component

$$s'_4 = \frac{\sigma}{c} F'_{4k} U'_k \equiv 0 \quad (s'_4 = ic\rho' = 0, \quad F'_{44} = 0).$$

But in the reference frame relative to which the medium moves

$$s_1 = \frac{\sigma}{c} F_{1k} U_k = \frac{\sigma}{c} (F_{12} U_2 + F_{13} U_3 + F_{14} U_4) = \frac{\sigma}{c} (-iE_x) ic\Gamma = \sigma\Gamma E_x,$$

$$s_2 = \frac{\sigma}{c} F_{2k} U_k = \frac{\sigma}{c} (F_{21} U_1 + F_{24} U_4) = \frac{\sigma}{c} (-cB_3 \Gamma V + (-iE_y) ic\Gamma) = \sigma\Gamma \{E_y + [\boldsymbol{VB}]_y\},$$

$$s_3 = \sigma\Gamma \{E_z + [\boldsymbol{VB}]_z\},$$

since

$$U_1 = \Gamma V, \quad U_2 = U_3 = 0, \quad U_4 = ic\Gamma.$$

The final result is obvious:

$$\boldsymbol{j} = \sigma\Gamma (\boldsymbol{E} + [\boldsymbol{VB}]). \tag{6.82}$$

Its meaning is quite clear: the current density in the medium with a conductivity σ is determined by the magnitude of an electric field in this medium; in accordance with Eq. (6.41) the magnitude of an electric field makes its appearance as a factor by σ in the case of $\Gamma \approx 1$.

The fourth equation defines the charge density associated with the conductivity current:

$$s_4 = ic\rho_{cond} = \frac{\sigma}{c} F_{4k} U_k = \frac{\sigma}{c} (i\Gamma \boldsymbol{EV}) = i\Gamma \left(\frac{jV}{c}\right), \tag{6.83}$$

or

$$\rho_{cond} = \Gamma \frac{jV}{c^2}$$

in complete agreement with Eq. (6.24).

It is worthwhile to consider Ohm's law in the case of moving media, i.e. the material equation (6.70). We shall see that the convection current $\rho \boldsymbol{v}$ and the conductivity current are closely interlocked as it becomes obvious right after j and $ic\rho$ are combined into a single 4-vector. The difference between a convection current and a conduction current is caused by the choice of a reference frame. Therefore it is natural that both currents alike induce a magnetic field.

We shall assume that a conductivity current represents a motion of charges with respect to a medium whereas a convection current arises due to the presence of charges in a medium owing to the motion of this medium.

Suppose that in a certain frame K' there is a conductivity current $j' = \sigma \boldsymbol{E}'$ and, besides, a charge density ρ'. These quantities constitute jointly a 4-current which can be transformed to any reference frame by means of Eq. (6.15a). Having expressed the j components and the density ρ through j' and ρ' in the reference frame K', we obtain

$$j_x = \Gamma (j'_x + V\rho'), \quad j_y = j'_y, \quad j_z = j'_z, \quad \rho = \Gamma \left(\rho' + \frac{V}{c^2} j'_x\right). \tag{6.84}$$

It is seen from the first formula of (6.84) that a conductivity current j_x incorporates a convection current $\Gamma V \rho' = V\rho$ so that it is not proportional to σ any more. It is inconvenient because at $\sigma = 0$ a conductivity current must turn into zero. How to distinguish a conductivity current in the general case? To do this, one must recall that if in the frame K' there is a charge density $\rho' = \rho_0$ we obtain a 4-current density in any other frame K (see Eq. (6.20))

$$s_i^{conv} = \rho_0 u_i, \tag{6.85}$$

where u_i is a 4-velocity of the charge. This current should be called a convection current; accordingly, s_i in Eq. (6.85) is supplemented with a superscript "conv". Suppose, we have a 4-current whose components are s_i, and we want to represent it as the sum of a conductivity current and a convection current. First of all, let us express ρ_0 through $\vec{s}$ and $\vec{V}$. Having multiplied both

sides of the equation $s_i = \rho_0 u_i$ by the corresponding components u_i and added up, we get $s_k u_k = \rho_0 u_k^2$; but according to Eq. (5.7) $u_k^2 = -c^2$, so that

$$\rho_0 = -\frac{s_k u_k}{c^2}. \tag{6.86}$$

Consequently, the convection current can be put down as

$$s_i^{conv} = -\frac{s_k u_k}{c^2} u_i. \tag{6.87}$$

In order to obtain components of a 4-conductivity current, one has to subtract components of Eq. (6.87) from s_i:

$$s_i^{cond} = s_i - s_i^{conv} = s_i + \frac{s_k u_k}{c^2} u_i. \tag{6.88}$$

On the other hand, in accordance with Eq. (6.81) the quantity s_i^{cond} can be put down as

$$s_i^{cond} = \frac{\sigma}{c} F_{ik} u_k. \tag{6.89}$$

Having equated these expressions, we obtain:

$$s_i + \frac{s_k u_k}{c^2} u_i = \frac{\sigma}{c} F_{ik} u_k. \tag{6.90}$$

Utilizing the definitions $\vec{s}(j_x, j_y, j_z, ic\rho)$, $\vec{V}(\gamma v_x, \gamma v_y, \gamma v_z, ic\gamma)$, we get

$$\boldsymbol{j} + \gamma^2 \boldsymbol{v}\left(\frac{\boldsymbol{jv}}{c^2} - \rho\right) = \sigma\gamma\,\{\boldsymbol{E} + [\boldsymbol{vB}]\} \tag{6.91}$$

in a three-dimensional form.

Let us separate the terms proportional to a conductivity σ in Eq. (6.91). For this purpose the left-hand and right-hand sides of Eq. (6.91) are multiplied by $\boldsymbol{v}$. Introducing the usual designations γ and β, we get

$$\frac{\boldsymbol{jv}}{c^2} = \rho\beta^2 + \frac{\sigma(\boldsymbol{Ev})}{c^2}\frac{1}{\gamma},$$

or

$$\frac{\boldsymbol{jv}}{c^2} - \rho = -\frac{\rho}{\gamma^2} + \frac{\sigma(\boldsymbol{Ev})}{c^2}\frac{1}{\gamma}. \tag{6.92}$$

Substituting Eq. (6.92) into Eq. (6.91), we finally obtain

$$\boldsymbol{j} = \rho\boldsymbol{v} + \sigma\gamma\left\{\boldsymbol{E} + [\boldsymbol{VB}] - \frac{\boldsymbol{v}}{c^2}(\boldsymbol{Ev})\right\}. \tag{6.93}$$

Thus, the term "conductivity current" can be attributed to the quantity $\boldsymbol{j}^{cond} = \boldsymbol{j} - \rho\boldsymbol{v}$. A field existing in a substance moving relative to a given reference frame is often denoted by $\boldsymbol{E}^*$

$$\boldsymbol{E}^* = \boldsymbol{E} + [\boldsymbol{VB}]. \tag{6.94}$$

Then Eq. (6.93) can be rewritten in the following form:

$$j^{cond} = \sigma\gamma\left\{E^{*} - \frac{v}{c^2}(E^{*}v)\right\}. \tag{6.95}$$

This equation resembles very much the force transformation equation (5.35). To complete the transition to equations of moving media electrodynamics, one has to find out how to describe boundary conditions when a media interface moves. The continuity condition for normal components of induction follows from the equations $\operatorname{div} D = 0$ and $\operatorname{div} B = 0$, which, according to Eqs. (6.66) and (6.60), keep their appearance on transition from one inertial frame of reference to another. Therefore, at the interface

$$D_{n1} = D_{n2}, \quad B_{n1} = B_{n2}. \tag{6.96}$$

Let us examine now the boundary conditions for tangent components of field strengths. Considering first the reference frame K' co-moving with the interface, we obtain the continuity condition for tangent components of E' and H' in this frame. But in terms of the frame K relative to which the interface moves at the velocity u, the fields E and H take the following form (see Eqs. (6.41) and (6.43)):

$$E' = E + [uB], \quad H' = H - [uD]. \tag{6.97}$$

Let us draw a perpendicular n to the interface plane and denote the projection of the velocity u on this perpendicular by u_n. Let us find the projections of Eq (6.97) on the plane perpendicular to n. Remembering that $[nE] = [n, E_n + E_t] = [nE_t]$, we shall write the equation $E'_{t1} = E'_{t2}$ as $[nE'_{t1}] = [nE'_{t2}]$, i. e.

$$[nE_1] + [n[uB_1]] = [nE_2] + [n[uB_2]],$$

or

$$[n, E_2 - E_1] = u(n(B_1 - B_2)) + (B_2 - B_1)(un).$$

Since according to Eq. (6.96) $nB_1 = nB_2$, we finally obtain

$$[n, E_2 - E_1] = u_n(B_2 - B_1), \tag{6.98}$$

and similarly,

$$[n, H_2 - H_1] = -u_n(D_2 - D_1). \tag{6.99}$$

This equation, together with Eq. (6.96), constitutes the boundary conditions for field vectors.

§ 6.9. The transformation of electric and magnetic moments. If we combine an electric and a magnetic moment P and M into a single antisymmetric tensor (6.33) we can immediately write transformation equations for components of these quantities.

Let us denote a polarization and magnetization, determined in the reference frame co-moving with substance, by $\boldsymbol{P}^0$ and $\boldsymbol{M}^0$ respectively. Then an observer relative to whom the substance moves at the velocity V will get

$$M_x = M_x^0, \quad M_y = \Gamma (M_y^0 + VP_z^0), \qquad M_z = \Gamma (M_z^0 - VP_y^0),$$
$$P_x = P_x^0, \quad P_y = \Gamma\left(P_y^0 - \frac{V}{c^2} M_z^0\right), \quad P_z = \Gamma\left(P_z^0 + \frac{V}{c^2} M_y^0\right). \tag{6.100}$$

These equations clear up at once the relationship between the three-dimensional vectors $\boldsymbol{P}$ and $\boldsymbol{M}$ introduced earlier. Here one can repeat everything that was said about the relationship between electric and magnetic fields. As a rule, magnetization is always accompanied with polarization, and vice versa. $\boldsymbol{P}$ or $\boldsymbol{M}$ can be equal to zero only in a specially chosen coordinate system. A polarized but not magnetized object is both polarized and magnetized in terms of an observer relative to whom this object moves. Indeed, suppose that in the frame K' in which the object is at rest

$$\boldsymbol{M}^0 = 0, \quad \boldsymbol{P}^0 (P_x^0, \ P_y^0, \ P_z^0) \neq 0.$$

Then in the frame K relative to which the object moves at the velocity V,

$$P_x = P_x^0, \qquad P_y = \Gamma P_y^0, \qquad P_z = \Gamma P_z^0,$$
$$M_x = 0, \qquad M_y = \Gamma V P_z^0, \quad M_z = -\Gamma V P_y^0.$$

Consequently in the frame K magnetization of the object will be observed. If the object moves at a non-relativistic velocity, i.e. $V/c \ll 1$ and $\Gamma \approx 1$,

$$\boldsymbol{P} = \boldsymbol{P}^0, \quad \text{and} \quad \boldsymbol{M} = [\boldsymbol{P}^0 \boldsymbol{V}].$$

This effect was found in experiments performed by Eichenwald (see [13], [29]). On the contrary, if in the frame K', relative to which the object is at rest,

$$\boldsymbol{P}^0 = 0, \quad \boldsymbol{M}^0 (M_x^0, \quad M_y^0, \quad M_z^0) \neq 0,$$

then in the frame K relative to which the object moves at the velocity V

$$M_x = M_x^0, \quad M_y = \Gamma M_y^0, \qquad M_z = \Gamma M_z^0,$$
$$P_x = 0, \qquad P_y = -\Gamma \frac{V}{c^2} M_z^0, \quad P_z = \Gamma \frac{V}{c^2} M_y^0. \tag{6.101}$$

Consequently, the object turns out to be polarized in the frame K. If the object moves at a non-relativistic velocity, then

$$M=M^0,\quad P=-\left[M^0\frac{V}{c^2}\right].$$

This implies, for example, that a moving permanent magnet carries an electric moment giving rise to the phenomenon of homopolar induction utilized in electrical engineering.

Here is an example illustrating these conclusions. Let the density of a current flowing along the rectangular loop $ABCD$ be equal to j, and the loop itself move at the velocity V relative to the frame K. Let us fix the frame K' to the loop (Fig. 6.5). In accordance with Eq. (6.24) a charge $\rho > 0$ appears in the section BC and a corresponding charge $\rho < 0$ in the section AD. It is obvious that the total charge appearing in the loop $ABCD$ is equal to zero. At the same time this loop possesses an electric moment directed along the y axis. We shall show that an elementary calculation coincides with conclusions of the STR.

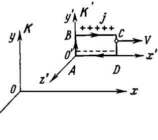


Fig. 6.5. Emergence of a dipole moment of a current loop when considered in the reference frame K relative to which this loop moves.

Although there is no dipole moment in K' and only the zth component of $\boldsymbol{M}$ is present, $P_y = -\Gamma\frac{V}{c^2}M_z^0$ emerges in K according to Eq. (6.101). In the frame K' the rectangular current $ABCD$ possesses the magnetic moment $\boldsymbol{jS}$ where the vector $\boldsymbol{S}$ is directed toward the negative z axis and is equal to ab (a and b being the sides of the rectangular loop). Assuming for the sake of simplicity the cross-section of the conductor to be equal to unity, we obtain $M_z^0 = -j^0ab$. The electric dipole moment emerging in the loop is not difficult to calculate. According to Eq. (6.24) $\rho=\Gamma\frac{V}{c^2}j_x^0$, the distance between BC and AD is equal to b, and the total charge in these sections is equal to ρa. The direction of this dipole moment coincides with that of the y axis. Therefore, $P_y=\rho ab=\Gamma\frac{V}{c^2}abj_x^0=-\frac{V}{c^2}\Gamma M_z^0$, just as it should be.

§ 6.10. Some problems involving the transformation of an electromagnetic field. *The field of a uniformly moving charge.* The magnetic and electric fields of a uniformly moving charge are most easily obtained by transformation of the fields existing in the frame K' in which the charge is at rest. In the case of a point

electric charge e resting in the frame K' we face an electrostatic problem since such a charge produces only an electric field. However, when the same charge is considered in terms of the frame K moving at the velocity $-V$ relative to K', it is found to generate a rectilinear current. A magnetic field induced by a rectilinear current is very well known: lines of force of such field form circles whose centres coincide with the current; the planes of these circles are perpendicular to the current direction. Naturally, these results follow from the field transformation equations.

Now, let a point charge be located at the origin of the frame K'. Then in this frame

$$\boldsymbol{B}'=0, \quad \boldsymbol{E}'=\frac{e}{4\pi\varepsilon}\frac{\boldsymbol{r}'}{r'^3},$$

or, when expressed in projections on the coordinate axes,

$$B'_x=0, \qquad B'_y=0, \qquad B'_z=0,$$

$$E'_x=\frac{e}{4\pi\varepsilon}\frac{x'}{r'^3}, \quad E'_y=\frac{e}{4\pi\varepsilon}\frac{y'}{r'^3}, \quad E'_z=\frac{e}{4\pi\varepsilon}\frac{z'}{r'^3},$$

where $r'^2=x'^2+y'^2+z'^2$. According to Eq. (6.36) we obtain in the frame K

$$E_x=E'_x, \qquad E_y=\Gamma E'_y, \qquad E_z=\Gamma E'_z, \tag{6.102}$$

$$B_x=B'_x=0, \quad B_y=-\Gamma\frac{V}{c^2}E'_z, \quad B_z=\Gamma\frac{V}{c^2}E'_y. \tag{6.103}$$

As $B_x=0$, the magnetic field in the frame K is located in the planes perpendicular to the x axis, i.e. in the planes perpendicular to the current direction. The equations describing the lines of force of the magnetic field take the following form in the frame K:

$$\frac{dy}{B_y}=\frac{dz}{B_z}, \quad \text{or} \quad \frac{dy}{dz}=\frac{B_y}{B_z}.$$

But

$$\frac{B_y}{B_z}=-\frac{E'_z}{E'_y}=-\frac{z'}{y'}=-\frac{z}{y},$$

since $z'=z$ and $y'=y$ under the Lorentz transformation. Consequently, the differential equation for the lines of force takes the form $dy/dz=-z/y$, or $y\,dy+z\,dz=0$, i.e. $d(y^2+z^2)=0$. Hence, it is obvious that we have the equation of a circle $y^2+z^2=\text{const}$ in the capacity of a first integral. Consequently, the lines of force represent circles with centres located on the current axis.

Surely, one can transform not only fields, but potentials as well. In the frame K' the scalar potential is equal to

$$\varphi'=\frac{1}{4\pi\varepsilon}\frac{e}{r'}=\frac{c}{i}\Phi'_4,$$

whereas the vector potential is equal to zero: $\boldsymbol{A}' = 0$. If the 4-potential $\vec{\Phi}$ has the components $\left(\boldsymbol{A}', \frac{i}{c}\varphi'\right)$ in the frame K', its components in the frame K take the following form in accordance with Eq. (6.14a):

$$\Phi_1 = \Gamma\left(\Phi_1' - i\frac{V}{c}\Phi_4'\right), \quad \Phi_2 = \Phi_2', \quad \Phi_3 = \Phi_3',$$
$$\Phi_4 = \Gamma\left(\Phi_4' + i\frac{V}{c}\Phi_1'\right).$$

Substituting the values of the 4-potential components in the frame K', we get

$$A_1 = \Gamma\left(-i\frac{V}{c}\frac{i}{c}\varphi'\right) = \Gamma\frac{V}{c^2}\varphi', \quad A_2 = 0, \quad A_3 = 0, \quad \frac{i}{c}\varphi = \Gamma\frac{i}{c}\varphi'.$$

Thus,

$$\boldsymbol{A} = \frac{\boldsymbol{V}}{c^2}\Gamma\varphi' = \frac{\boldsymbol{V}}{c^2}\varphi, \quad \varphi = \Gamma\varphi'. \tag{6.104}$$

Now we have to express r' entering into φ' via the charge coordinates in the frame K. According to the Lorentz transformation

$$x' = \Gamma(x - Vt), \quad y' = y, \quad z' = z, \tag{6.105}$$

and the expression for r'^2 will be written as

$$r'^2 = x'^2 + y'^2 + z'^2 = \Gamma^2(x - Vt)^2 + y^2 + z^2 =$$
$$= \Gamma^2\left[(x - Vt^2) + \frac{y^2 + z^2}{\Gamma^2}\right] = \Gamma^2\mathfrak{R}^2, \quad r' = \Gamma\mathfrak{R}, \tag{6.106}$$

where the following designation is introduced:

$$\mathfrak{R}^2 = (x - Vt)^2 + \left(1 - \frac{V^2}{c^2}\right)(y^2 + z^2). \tag{6.107}$$

Making use of Eq. (6.107), one can express the scalar potential φ, defined in Eq. (6.104), through $\mathfrak{R}$:

$$\varphi = \Gamma\varphi' = \Gamma\frac{1}{4\pi\varepsilon}\frac{e}{r'} = \frac{1}{4\pi\varepsilon}\frac{e}{\mathfrak{R}}.$$

Accordingly, the vector potential $\boldsymbol{A}$ can be written in the form

$$\boldsymbol{A} = \frac{\boldsymbol{V}}{c^2}\varphi = \frac{1}{4\pi\varepsilon}\frac{e\boldsymbol{V}}{c^2\mathfrak{R}}.$$

Let us rewrite the expressions for the $\boldsymbol{E}$ components of the field, taking into account Eqs. (6.102), (6.105) and (6.107). We

get:

$$E_x = \frac{e(x - Vt)}{\Gamma^2 4\pi\varepsilon \Re^3},$$
$$E_y = \frac{ey}{\Gamma^2 4\pi\varepsilon \Re^3}, \qquad (6.108)$$
$$E_z = \frac{ez}{\Gamma^2 4\pi\varepsilon \Re^3}.$$

In the frame K' the charge is located at the origin O' (i.e. at the point $x' = 0$). Its coordinates at the moment t in the frame K will be as follows: $x_0 = Vt$, $y_0 = 0$, $z_0 = 0$. Let us introduce one more vector, $\boldsymbol{R}$, directed from the point O', where the charge is located

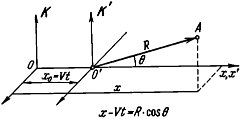


Fig. 6.6. To the calculation of an electric and a magnetic field of a uniformly moving charge.

toward the observation point A whose coordinates are (x, y, z) (Fig. 6.6). The vector $\boldsymbol{R}$ will take the form

$$\boldsymbol{R} = (x - Vt)\,\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}, \qquad (6.109)$$

where $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ are unit vectors along the x, y, z axes respectively. Having multiplied the components of Eq. (6.108) by $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ respectively, we obtain

$$\boldsymbol{E} = E_x\boldsymbol{i} + E_y\boldsymbol{j} + E_z\boldsymbol{k} = \frac{1}{4\pi\varepsilon}\frac{e}{\Gamma^2}\frac{\boldsymbol{R}}{\Re^3}.$$

If one introduces the angle θ between the charge motion direction (i.e. the x axis) and the radius vector $\boldsymbol{R}$, then

$$x - Vt = R\cos\theta, \quad R^2 = R^2\cos^2\theta + y^2 + z^2,$$

and consequently,

$$y^2 + z^2 = R^2\sin^2\theta. \qquad (6.110)$$

Taking into account Eqs. (6.109) and (6.110), one can rewrite Eq. (6.107) as

$$\Re^2 = R^2\left(1 - \frac{V^2}{c^2}\sin^2\theta\right),$$

whereupon the expression for $\boldsymbol{E}$ can be finally represented in the form

$$\boldsymbol{E}=\frac{1}{4\pi\varepsilon}\frac{e\boldsymbol{R}}{R^3}\frac{1-\frac{V^2}{c^2}}{\left(1-\frac{V^2}{c^2}\sin^2\theta\right)^{3/2}}. \tag{6.111}$$

Eq. (6.111) presents the electric field of a moving charge in very convenient variables, i.e. the distance R from the moving charge and the angle θ formed by the direction to the point at which the field is sought and the charge motion direction. Eq. (6.111) shows that the magnitude of the field depends on the angle θ. At a fixed R the minimum magnitude of the field corresponds to the charge motion direction ($\theta = 0, \pi$):

$$E_{\parallel}=\frac{1}{4\pi\varepsilon}\frac{e}{R^2}\left(1-\frac{V^2}{c^2}\right),$$

and the maximum magnitude of the field is observed in the direction perpendicular to the motion ($\theta = \pi/2$):

$$E_{\perp}=\frac{1}{4\pi\varepsilon}\frac{e}{R^2}\frac{1}{\sqrt{1-V^2/c^2}}.$$

A field strength magnitude depends on a charge motion velocity, with $E_{\parallel}$ decreasing and $E_{\perp}$ growing as the velocity increases. The electric field of a charge moving at a relativistic velocity is localized within two narrow solid angles whose boundary surface is approximately determined from the relation $(V^2\sin^2\theta/c^2)\approx 1$; the axial line of these solid angles is perpendicular to the charge motion direction.

The magnetic field $\boldsymbol{B}$ of a charge moving in the frame K can be found by means of Eq. (6.44) ($\boldsymbol{B}'=0$ in the frame K'):

$$\boldsymbol{B}=\frac{1}{c^2}[\boldsymbol{V}\boldsymbol{E}]. \tag{6.112}$$

When the velocity of the charge is low, the fields *in vacuo* are described by the following approximate relations:

$$\boldsymbol{E}=\frac{1}{4\pi\varepsilon_0}\frac{e\boldsymbol{R}}{R^3}$$

and

$$\boldsymbol{B}=\frac{1}{4\pi\varepsilon_0 c^2}\frac{e[\boldsymbol{V}\boldsymbol{R}]}{R^3}=\frac{\mu_0}{4\pi}\frac{[e\boldsymbol{V}\boldsymbol{R}]}{R^3}. \tag{6.113}$$

Eq. (6.113) represents the *Biot-Savart law*.

The interaction of two moving charges. Let two charges e_1 and e_2 move in parallel at the same velocity V. Let us determine the interaction force between them in the reference frame K relative

to which they move. First, we shall find the force acting on the charge e_1.

The charge e_1 experiences the action of an electric and a magnetic field induced by the charge e_2. The force acting on the charge e_1 is the Lorentz force:

$$\boldsymbol{F}_1 = e_1 \{\boldsymbol{E}_2 + [\boldsymbol{V}\boldsymbol{B}_2]\}.$$

Taking into account Eq. (6.112), one can write down

$$\boldsymbol{F}_1 = e_1\boldsymbol{E}_2 + \frac{e_1}{c^2}[\boldsymbol{V}[\boldsymbol{V}\boldsymbol{E}_2]] = e_1\boldsymbol{E}_2 + \frac{e_1}{c^2}\boldsymbol{V}(\boldsymbol{V}\boldsymbol{E}_2) - \frac{e_1}{c^2}V^2\boldsymbol{E}_2 =$$
$$= e_1\left(1 - \frac{V^2}{c^2}\right)\boldsymbol{E}_2 + \frac{e_1}{c^2}\boldsymbol{V}(\boldsymbol{V}\boldsymbol{E}_2). \qquad (6.114)$$

$\boldsymbol{E}_2$ can be found from Eq. (6.111) in which $\boldsymbol{R}$ is assumed to be the radius vector drawn from the charge e_2 to e_1 and θ the angle between $\boldsymbol{R}$ and the charge motion velocity direction $\boldsymbol{V}$. Substituting Eq. (6.111) into Eq. (6.114), we obtain

$$\boldsymbol{F}_1 = \frac{e_1e_2}{4\pi\varepsilon R^3}\frac{\left(1 - \frac{V^2}{c^2}\right)^2\boldsymbol{R}}{\left(1 - \frac{V^2}{c^2}\sin^2\theta\right)^{3/2}} + \frac{e_1e_2\boldsymbol{V}V\cos\theta\left(1 - \frac{V^2}{c^2}\right)}{4\pi\varepsilon c^2R^2\left(1 - \frac{V^2}{c^2}\sin^2\theta\right)^{3/2}} =$$
$$= \frac{e_1e_2}{4\pi\varepsilon R^2}\frac{1 - \frac{V^2}{c^2}}{\left(1 - \frac{V^2}{c^2}\sin^2\theta\right)^{3/2}}\left[\left(1 - \frac{V^2}{c^2}\right)\frac{\boldsymbol{R}}{R} + \frac{\boldsymbol{V}V}{c^2}\cos\theta\right],$$

whence the component along the motion direction is

$$F_x = \frac{e_1e_2}{4\pi\varepsilon R^2}\frac{\left(1 - \frac{V^2}{c^2}\right)\cos\theta}{\left(1 - \frac{V^2}{c^2}\sin^2\theta\right)^{3/2}},$$

and the component perpendicular to the motion direction

$$F_y = \frac{e_1e_2}{4\pi\varepsilon R^2}\frac{\left(1 - \frac{V^2}{c^2}\right)^2\sin\theta}{\left(1 - \frac{V^2}{c^2}\sin^2\theta\right)^{3/2}}. \qquad (6.115)$$

Let the charges be located on a straight line parallel to the y axis, with one of the charges being at the x axis, so that the distance between the charges is equal to y. Then $\theta = \pi/2$, $F_x = 0$, and

$$F_y = \frac{e_1e_2}{4\pi\varepsilon y^2}\sqrt{1 - \frac{V^2}{c^2}}. \qquad (6.116)$$

This equation can be obtained very simply. In the frame K' where the two charges are at rest, the interaction between them is

of electrostatic nature and the interaction force is equal to $e_1e_2/4\pi\varepsilon y^2$. The transformation of this force on transition from the frame K' to the frame K by means of Eq. (5.35) yields Eq. (6.116). According to Eq. (6.116) the charges repel each other in the frame K. But in the frame K the charges move producing two parallel currents flowing in the same direction. When such currents flow along conductors, they attract each other. There is no contradiction here since the physical situations are different. Let us consider the expression for a force (Eq. (6.116)) *in vacuo* in the case of non-relativistic velocities V:

$$F_y \approx \frac{e_1e_2}{4\pi\varepsilon_0 y^2}\left(1 - \frac{V^2}{2c^2} + \ldots\right) = \frac{e_1e_2}{4\pi\varepsilon_0 y^2} - \frac{e_1e_2}{4\pi y^2}\mu_0\frac{V^2}{2}. \qquad (6.117)$$

On the other hand, following Ampere, the interaction force between two current elements $e_1\boldsymbol{V}$ and $e_2\boldsymbol{V}$ *in vacuo* can be written as

$$F_{12} = \frac{\mu_0}{4\pi}\frac{[e_1\boldsymbol{V}\,[e_2\boldsymbol{V}, \boldsymbol{R}]]}{R^3} = \frac{\mu_0 e_1e_2}{4\pi}\frac{[\boldsymbol{V}\,[\boldsymbol{V}\boldsymbol{R}]]}{R^3} = -\frac{\mu_0 e_1e_2}{4\pi}V^2\frac{\boldsymbol{R}}{R^3}.$$

Taking into account that $\boldsymbol{R} = y\boldsymbol{j}$, we finally get

$$F_y = -\frac{\mu_0 e_1e_2}{4\pi y^2}V^2. \qquad (6.118)$$

The force defined by Eq. (6.117) and observed in the reference frame K relative to which the charges move, consists of the Coulomb repulsion and the Ampere attraction (with an accuracy of the factor 1/2). The force expressed in the form of Eq. (6.117) can be used for the explanation of current interactions in conductors only with certain stipulations. Neutral current-carrying conductors must attract each other in these conditions. However, a current-carrying conductor is neutral only in one reference frame (§ 6.1). That is why the Coulomb repulsion ought to be taken into consideration. For all that, it usually seems to be weaker than the attraction.

§ 6.11. An energy-momentum-tension tensor of an electromagnetic field *in vacuo*. A transition to four-dimensional quantities combines the quantities whose interrelationship was imperceptible in a three-dimensional approach. In the case of a free particle one 4-vector combines energy and momentum. An electric and a magnetic field constitute an electromagnetic field tensor in 4-space. An energy and a momentum of an electromagnetic field turn out to be components of a tensor which, apart from an energy (a scalar in a three-dimensional case) and a momentum (a three-dimensional vector), comprises also a three-dimensional-tension tensor of Maxwell. Here we shall have to quote the results of the Maxwell theory in a three-dimensional form.

1. *The energy conservation law for charges and a field.* This law follows directly from Maxwell's equations: multiplying Eq. (6.56a)

scalarwise by $\boldsymbol{E}$ and Eq. (6.57a) by $\boldsymbol{H}$ and subtracting the expressions thus obtained, we get

$$\boldsymbol{H}\operatorname{rot}\boldsymbol{E}-\boldsymbol{E}\operatorname{rot}\boldsymbol{H}=-\boldsymbol{jE}-\dot{\boldsymbol{D}}\boldsymbol{E}-\dot{\boldsymbol{B}}\boldsymbol{H}.$$

Making use of the following identities $\boldsymbol{H}\operatorname{rot}\boldsymbol{E}-\boldsymbol{E}\operatorname{rot}\boldsymbol{H}=\operatorname{div}[\boldsymbol{EH}]$ and $(d/dt)(\boldsymbol{ED}+\boldsymbol{BH})=2(\dot{\boldsymbol{D}}\boldsymbol{E}+\dot{\boldsymbol{B}}\boldsymbol{H})$ (the last one is valid for an isotropic medium in which $\boldsymbol{D}=\varepsilon\boldsymbol{E}$, $\boldsymbol{B}=\mu\boldsymbol{H}$), we get

$$\frac{d}{dt}\left(\frac{\boldsymbol{ED}+\boldsymbol{BH}}{2}\right)=-\boldsymbol{jE}-\operatorname{div}[\boldsymbol{EH}],$$

whence, after integration over an arbitrary volume $\mathscr{V}$ and using the Gauss-Ostrogradsky theorem, we come to

$$\frac{dW}{dt}=-\int_{\mathscr{V}}\boldsymbol{jE}\,d\mathscr{V}-\oint_{S}\boldsymbol{S}\,d\boldsymbol{S}. \tag{6.119}$$

The left-hand side of Eq. (6.119) represents a time variation of an electromagnetic field energy in a volume $\mathscr{V}$. This energy is defined in the Maxwell theory via an energy density (an energy per unit of volume):

$$w=\frac{\boldsymbol{ED}+\boldsymbol{BH}}{2} \tag{6.120}$$

by integrating over a vol'ıme:

$$W=\int_{\mathscr{V}}w\,d\mathscr{V}. \tag{6.121}$$

Let us consider the simplest case of charges *in vacuo*. In this case $\boldsymbol{j}=\rho\boldsymbol{v}$, while the force density acting on these charges, i.e. a Lorentz force density, is

$$\boldsymbol{f}^L=\rho(\boldsymbol{E}+[\boldsymbol{vB}])=\rho\boldsymbol{E}+[\boldsymbol{jB}]. \tag{6.122}$$

This force is introduced into the theory in order to correlate the field theory with the field of force acting on charged objects located in the field. Eq. (6.119) features an expression $\boldsymbol{jE}$. Multiplying Eq. (6.122) scalarwise by $\boldsymbol{v}$, we obtain $\boldsymbol{f}^L\boldsymbol{v}=\rho\boldsymbol{Ev}=\boldsymbol{jE}$. Therefore, one of the terms of the right-hand side represents a work performed by a field on a charge in this case. In accordance with the energy conservation law this work must turn into a kinetic energy of particles T. Consequently,

$$\int_{\mathscr{V}}\boldsymbol{jE}\,d\mathscr{V}=\int\boldsymbol{f}^L\boldsymbol{v}\,d\mathscr{V}=\frac{dT}{dt}. \tag{6.123}$$

The second term of the right-hand side of Eq. (6.119) represents the Poynting vector

$$\boldsymbol{S}=[\boldsymbol{EH}], \tag{6.124}$$

while the integral itself represents a flux of the vector $\boldsymbol{S}$ through the surface enclosing the volume $\mathcal{V}$. The integrand also includes $d\boldsymbol{S} = \boldsymbol{n}dS$, a product of a surface area dS and a unit vector $\boldsymbol{n}$ of its normal. Hence, the energy conservation law for charges and a field can be written down as follows:

$$\frac{d}{dt}(T+W) = -\oint_S \boldsymbol{S}\, d\boldsymbol{S}. \qquad (6.125)$$

The Poynting vector (6.124) is usually interpreted as an energy flux per unit time through a unit area oriented normally to the Poynting vector. Such an interpretation does not necessarily follow from the Maxwell equations. The direct consequence of the Maxwell equations is the integral relation (6.119) which can be regarded as the energy conservation law. It is clear that any $\boldsymbol{S}'$ addition to the Poynting vector $\boldsymbol{S}$, satisfying the condition div $\boldsymbol{S}'=0$, does not vary the relation (6.119). The generally accepted interpretation, however, is confirmed by experiment.

2. *The momentum conservation law for charges and a field.* The momentum conservation law can be treated as follows. Multiplying Eq. (6.56a) vectorwise by $\boldsymbol{B}$ and Eq. (6.57a) by $\boldsymbol{D}$, and adding the equations obtained termwise, we get

$$\mu[\boldsymbol{H}\operatorname{rot}\boldsymbol{H}] + \varepsilon[\boldsymbol{E}\operatorname{rot}\boldsymbol{E}] = -[\boldsymbol{j}\boldsymbol{B}] + \varepsilon\mu[\boldsymbol{H}\dot{\boldsymbol{E}}] - \varepsilon\mu[\boldsymbol{E}\dot{\boldsymbol{H}}].$$

We have taken into account that $\boldsymbol{D} = \varepsilon\boldsymbol{E}$ and $\boldsymbol{B} = \mu\boldsymbol{H}$; finally:

$$\mu[\boldsymbol{H}\operatorname{rot}\boldsymbol{H}] + \varepsilon[\boldsymbol{E}\operatorname{rot}\boldsymbol{E}] = -[\boldsymbol{j}\boldsymbol{B}] - \varepsilon\mu\frac{d}{dt}[\boldsymbol{E}\boldsymbol{H}]. \qquad (6.126)$$

Let us make use of the vector identity

$$\boldsymbol{a}\operatorname{div}\boldsymbol{a} - [\boldsymbol{a}\operatorname{rot}\boldsymbol{a}] = \frac{\partial}{\partial x_\alpha}\left(a_\alpha a_\beta - \frac{1}{2}\delta_{\alpha\beta}a_\gamma^2\right)\boldsymbol{m}_\beta.$$

Subtracting the left-hand and right-hand sides of Eq. (6.126) respectively from the identity

$$\mu\boldsymbol{H}\operatorname{div}\boldsymbol{H} + \varepsilon\boldsymbol{E}\operatorname{div}\boldsymbol{E} \equiv \rho\boldsymbol{E}$$

(see Eqs. (6.56b) and (6.57b)), we get the final equation

$$\frac{\partial}{\partial x_\alpha}\left(\varepsilon E_\alpha E_\beta + \mu H_\alpha H_\beta + \delta_{\alpha\beta}\frac{\varepsilon E^2 + \mu H^2}{2}\right)\boldsymbol{m}_\beta =$$
$$= \rho\boldsymbol{E} + [\boldsymbol{j}\boldsymbol{B}] + \varepsilon\mu\frac{d}{dt}[\boldsymbol{E}\boldsymbol{H}],$$

which can be rewritten in the form

$$\frac{\partial T_{\alpha\beta}}{\partial x_\alpha}\boldsymbol{m}_\beta = \boldsymbol{f}^L + \frac{d}{dt}[\boldsymbol{D}\boldsymbol{B}]. \qquad (6.127)$$

where the tension tensor of Maxwell is introduced

$$T_{\alpha\beta} = \varepsilon E_\alpha E_\beta + \mu H_\alpha H_\beta - \delta_{\alpha\beta} w = E_\alpha D_\beta + H_\alpha B_\beta - \delta_{\alpha\beta} w. \qquad (6.128)$$

The tensor (6.128), being symmetric *in vacuo* and isotropic media, is asymmetric in anisotropic media where it is defined according to the last equation of (6.128). Integrating Eq. (6.127) over an arbitrary volume in the region where an electromagnetic field exists, we get

$$\int_{\mathscr{V}} \frac{\partial T_{\alpha\beta}}{\partial x_\alpha} \boldsymbol{m}_\beta \, d\mathscr{V} = \int_{\mathscr{V}} \boldsymbol{f}^L \, d\mathscr{V} + \frac{d}{dt} \int_{\mathscr{V}} [\boldsymbol{DB}] \, d\mathscr{V}. \qquad (6.129)$$

It was again assumed that in Eq. (6.127) we deal with free charges *in vacuo*, subjected to the Lorentz force (6.122).

According to the second law of Newton

$$\int_{\mathscr{V}} \boldsymbol{f} \, d\mathscr{V} = \frac{d\boldsymbol{P}}{dt}, \qquad (6.130)$$

where $\boldsymbol{P}$ is the momentum of charges enclosed within the volume $\mathscr{V}$. The integral with respect to volume entering into the left-hand side of Eq. (6.129) transforms into the integral with respect to the surface enveloping the volume $\mathscr{V}$:

$$\int_{\mathscr{V}} \frac{\partial T_{\alpha\beta}}{\partial x_\alpha} \boldsymbol{m}_\beta \, d\mathscr{V} = \oint_S T_{\alpha\beta} n_\alpha \boldsymbol{m}_\beta \, dS. \qquad (6.131)$$

The expressions

$$T_{\alpha\beta} n_\alpha \boldsymbol{m}_\beta \, dS$$

represent a force acting on an infinitesimal surface area dS whose normal's components are n_α. The vectors $\boldsymbol{m}_\beta$ are unit vectors of the Cartesian coordinate system. We could already write down the momentum conservation law if we knew what should be regarded as a momentum of an electromagnetic field in matter. So let us confine ourselves to the case of vacuum where $[\boldsymbol{DB}] = (1/c^2)[\boldsymbol{EH}] = \boldsymbol{S}/c^2$. Then taking into account Eqs. (6.130) and (6.131), Eq. (6.129) can be rewritten in the form

$$\frac{d}{dt}(\boldsymbol{P} + \boldsymbol{G}) = \oint_S T_{\alpha\beta} n_\alpha \boldsymbol{m}_\beta \, dS, \qquad (6.132)$$

having defined a field momentum density g *in vacuo* as

$$g = S/c^2 \tag{6.133}$$

and consequently a field momentum in the volume $\mathcal{V}$ as

$$G = \int_{\mathcal{V}} g \, d\mathcal{V}.$$

Equation (6.132) and the definition (6.133) express the momentum conservation law. For a complete field, when on the boundary surface $T_{\alpha\beta} = 0$, we obtain the conservation law $(d/dt)(P + G) = 0$. The tensor $T_{\alpha\beta}$ is not defined unambiguously: the Maxwell equations yield only the integral equation (6.132), and if one adds a component of an arbitrary tensor $T'_{\alpha\beta}$, satisfying the condition $(\partial T'_{\alpha\beta}/\partial x_\alpha) = 0$, to each component of the vector G,

$$\oint T'_{\alpha\beta} n_\alpha m_\beta \, dS = \int_{\mathcal{V}} \frac{\partial T'_{\alpha\beta}}{\partial x_\alpha} m_\beta \, d\mathcal{V} = 0,$$

and Eq. (6.132) retains its validity all the same. Here we proceed in much the same way as we did when selecting an expression for the Poynting vector from the energy conservation theorem or finding an expression for a displacement current density. Our selection depends on the correctness of all of its consequences. The tension tensor (6.128) *in vacuo* where $D = \varepsilon_0 E$, $B = \mu_0 H$, together with the definition of a momentum (Eq. (6.133)), yields reasonable physical results.

In conclusion note that Eq. (6.132) makes it clear that the definitions of a momentum density and a tension tensor are closely interrelated. Having redefined a definition for a momentum density, we modify at once an expression of $T_{\alpha\beta}$ (see § 6.12).

Let us summarize the results that we obtained for the case of vacuum: as a consequence of the Maxwell equations, the momentum density defined by Eq. (6.133) ought to be assigned to an electromagnetic field *in vacuo*. Then Eq. (6.132) expresses the Newton law: an increment of the total momentum of charges and of a field in a volume $\mathcal{V}$ is equal to the sum of forces acting upon this volume. These forces can be written down in the form of surface forces, i.e. the forces acting on a surface enveloping the volume $\mathcal{V}$.

A transition to four-dimensional terms can be accomplished as follows. First, let us prove that a 4-force density $\vec{f}$ (see Eq. (6.54)) can be rewritten as a four-dimensional divergence of a tensor T_{ik}:

$$f_i = \frac{1}{c} F_{ik} s_k = \frac{\partial T_{ik}}{\partial x_k},$$

where T_{ik} is an energy-momentum-tension tensor *; the components of this tensor have the following form:

$$T_{ik}=\frac{1}{c}F_{im}f_{mk}+\frac{1}{4c}\delta_{ik}(F_{sn}f_{sn}). \qquad (6.134)$$

In the first term of the right-hand side of Eq. (6.134) summation is performed over m, and in the second one over s and n; F_{im} and f_{sn} are the corresponding components of tensors (6.29a) and (6.31).

In order to obtain Eq. (6.134) we shall need the Maxwell equations expressed as Eqs. (6.60) and (6.67); here we shall rewrite them in a more convenient form:

$$\frac{\partial f_{kl}}{\partial x_l}=s_k, \qquad (6.135)$$

$$\frac{\partial F_{lk}}{\partial x_l}+\frac{\partial F_{li}}{\partial x_k}=-\frac{\partial F_{kl}}{\partial x_l}. \qquad (6.136)$$

Let us transform the four-dimensional force density:

$$f_i=\frac{1}{c}F_{ik}s_k=\frac{1}{c}F_{ik}\frac{\partial f_{kl}}{\partial x_l}=\frac{1}{c}\left\{\frac{\partial}{\partial x_l}(F_{ik}f_{kl})-f_{kl}\frac{\partial F_{ik}}{\partial x_l}\right\}. \qquad (6.137)$$

Here we made use of Eq. (6.135) and applied the rule of a product differentiation.

Now we shall deal with the second term in the last link of Eq. (6.137):

$$f_{kl}\frac{\partial F_{ik}}{\partial x_l}=\frac{1}{2}\left(f_{kl}\frac{\partial F_{ik}}{\partial x_l}+f_{lk}\frac{\partial F_{kl}}{\partial x_l}\right)=$$

$$=\frac{1}{2}\left(f_{kl}\frac{\partial F_{lk}}{\partial x_l}+f_{kl}\frac{\partial F_{li}}{\partial x_k}\right)=\frac{1}{2}f_{kl}\left(\frac{\partial F_{lk}}{\partial x_l}+\frac{\partial F_{li}}{\partial x_k}\right)=$$

$$=-\frac{1}{2}f_{kl}\frac{\partial F_{kl}}{\partial x_i}=-\frac{1}{2}\sqrt{\frac{\mu_0}{e_0}}f_{kl}\frac{\partial f_{kl}}{\partial x_i}=-\frac{1}{4}\sqrt{\frac{\mu_0}{e_0}}\frac{\partial}{\partial x_i}(f_{kl})^2=$$

$$=-\frac{1}{4}\frac{\partial}{\partial x_i}(f_{sn}F_{sn}). \qquad (6.138)$$

The following operations are carried out in the equation chain (6.138). The transition to the second link is based on antisymmetry of the tensors f_{kl} and F_{ik} permitting of exchanging indices in each term of the product without changing the product in the process·

$$f_{kl}\frac{\partial F_{ik}}{\partial x_l}=f_{lk}\frac{\partial F_{kl}}{\partial x_l}. \qquad (6.139)$$

* We shall denote the four-dimensional tensor (6.134) by the same letter T that we used in the case of the three-dimensional tensor (6.128); this should not lead to any misunderstanding because, as it will turn out, nine components of (6.134) for i and k changing from 1 to 3 coincide with (6.128).

Consequently, instead of one term we take a half-sum of two equal expressions given by the left-hand and right-hand sides of Eq. (6.139). The third link involves a substitution of mute indices which does not change a summation result: the index l is replaced by k and vice versa, i.e. $f_{lk}\frac{\partial F_{kl}}{\partial x_l}$ is replaced by $f_{kl}\frac{\partial F_{ll}}{\partial x_k}$. The common factor f_{kl} is taken out of the brackets in the fourth link, while in the fifth link Eq. (6.136) is used. In the sixth link of the equation we replaced F_{kl} in accordance with Eq. (6.58); this operation is valid only for vacuum.

Eq. (6.138), however, remains valid also for a uniform isotropic medium. It can be easily shown that the components of tensors $\mathfrak{f}$ and $\mathfrak{F}$ are proportional as before in such a medium although spatial and temporal components have different proportionality factors. From the general appearance of the tensors $\mathfrak{F} = (c\boldsymbol{B}, -i\boldsymbol{E})$ and $\mathfrak{f} = (\boldsymbol{H}, -ic\boldsymbol{D})$ it is clear that the spatial components tie together the vectors $\boldsymbol{B}$ and $\boldsymbol{H}$, while the temporal ones the vectors $\boldsymbol{D}$ and $\boldsymbol{E}$. In order to get the necessary relations $\boldsymbol{D} = \varepsilon\boldsymbol{E}$ and $\boldsymbol{B} = \mu\boldsymbol{H}$, one has to assume

$$f_{\alpha\beta} = aF_{\alpha\beta}, \quad a = \frac{\mu_0}{\mu}\sqrt{\frac{\varepsilon_0}{\mu_0}} = \frac{1}{\mu c}, \tag{6.140}$$

$$f_{k4} = bF_{k4}, \quad b = \frac{\varepsilon}{\varepsilon_0}\sqrt{\frac{\varepsilon_0}{\mu_0}} = \varepsilon c. \tag{6.141}$$

But then, starting from the fifth link of Eq. (6.138), the subsequent chain of equations will be rewritten as follows:

$$\begin{aligned}\frac{1}{2} f_{kl}\frac{\partial F_{kl}}{\partial x_l} &= \frac{1}{2}a^{-1}f_{\alpha\beta}\frac{\partial f_{\alpha\beta}}{\partial x_l} + \frac{1}{2}b^{-1}f_{k4}\frac{\partial f_{k4}}{\partial x_l} = \frac{a^{-1}}{4}\frac{\partial}{\partial x_l}(f_{\alpha\beta})^2 + \\ &+ \frac{b^{-1}}{4}\frac{\partial}{\partial x_l}(f_{k4})^2 = \frac{1}{4}\frac{\partial}{\partial x_l}(f_{\alpha\beta}F_{\alpha\beta}) + \frac{1}{4}\frac{\partial}{\partial x_l}(f_{k4}F_{k4}) = \\ &= \frac{1}{4}\frac{\partial}{\partial x_l}(f_{kl}F_{kl}) = \frac{1}{4}\frac{\partial}{\partial x_l}(f_{sn}F_{sn}).\end{aligned} \tag{6.142}$$

The last link of Eq. (6.138) and the third link of Eq. (6.142) are written according to the rule of product differentiation:

$$f_{kl}\frac{\partial f_{k'}}{\partial x_l} = \frac{1}{2}\frac{\partial}{\partial x_l}(f_{kl})^2.$$

Since the factors a and b in Eqs. (6.140) and (6.141) are constant, they can be brought under the differential symbol. And finally, for the sake of convenience some other mute indices are introduced in summation; in so doing, we are not changing the sum $f_{kl}F_{kl} = f_{sn}F_{sn}$. Therefore, the results obtained for vacuum and a uniform isotropic medium prove to be the same.

Of course, this result is formally obvious because in the SI system vacuum is just one of uniform and isotropic media as long as only the relations $\boldsymbol{D} = \varepsilon \boldsymbol{E}$ and $\boldsymbol{B} = \mu \boldsymbol{H}$ are essential.

The first term of Eq. (6.137) can be combined with the second term of Eq. (6.138) or Eq. (6.142), given in a final form, provided both terms are differentiated with respect to the same variables. However, a differentiation with respect to another variable can be performed by means of the Kronecker delta

$$\frac{\partial}{\partial x_l} = \delta_{ll} \frac{\partial}{\partial x_l}.$$

Now we can write down the expression for f_l in a complete form (see Eq. (6.137)):

$$f_l = \frac{1}{c}\left\{\frac{\partial}{\partial x_l}(F_{lk} f_{kl}) + \frac{1}{4}\frac{\partial}{\partial x_l}(f_{sn}F_{sn})\right\} =$$
$$= \frac{1}{c}\frac{\partial}{\partial x_l}\left\{F_{lk}F_{kl} + \frac{1}{4}\delta_{ll}(f_{sn}F_{sn})\right\}. \qquad (6.143)$$

Let us substitute the mute summation indices in the first summand: the index m will replace k, and the index k will replace l in the first and second summands. Then we finally obtain

$$f_l = \frac{\partial}{\partial x_k}\left\{\frac{1}{c}F_{lm}f_{mk} + \frac{1}{4c}\delta_{lk}(f_{sn}F_{sn})\right\} \equiv \frac{\partial T_{lk}}{\partial x_k}, \qquad (6.144)$$

where T_{lk} is defined according to Eq. (6.134).

Thus, the components of a 4-force $\vec{f}$ can be expressed through the components of a tensor T_{lk} depending on the field vectors $\boldsymbol{E}$ and $\boldsymbol{B}$. Recall that the components of tensors $\mathfrak{F}$ and $\mathfrak{f}$ are proportional *in vacuo* according to Eq. (6.58), the proportionality coefficient being the same.

Owing to this circumstance and the definition of the tensor T_{lk} (Eq. (6.134)), one can see it is symmetric *in vacuo*, i.e. $T_{lk} = T_{kl}$. It implies that this tensor has ten independent components. In the case of matter a 4-tensor loses its symmetric properties.

Let us find now the T_{lk} components expressed through electromagnetic field vectors. Consider first the expression $f_{sn}F_{sn}$. This is precisely the sum of pairwise products of the respective components of matrices (6.29a) and (6.31). The requisite components are immediately seen from the definition of tensors $\mathfrak{f}(\boldsymbol{H}, -ic\boldsymbol{D})$ and $\mathfrak{F}(c\boldsymbol{B}, -i\boldsymbol{E})$. Designating the coefficient in δ_{lk} as Λ, we find

$$\Lambda = \frac{1}{4c} f_{sn}F_{sn} = \frac{1}{4c}\cdot 2(c\boldsymbol{BH} - c\boldsymbol{DE}) = \frac{\boldsymbol{BH}}{2} - \frac{\boldsymbol{DE}}{2}. \qquad (6.145)$$

The digit 2 appeares in front of the parenthesis because due to antisymmetric properties of f and F the product of pairwise com-

ponents yields the same expression $c(\boldsymbol{BH}-\boldsymbol{DE})$ twice. Now the equation of T_{ik} components can be rewritten as follows ($-f_{km}$ are substituted for f_{mk}):

$$T_{ik}=-\frac{1}{c}F_{im}f_{km}+\delta_{ik}\Lambda. \tag{6.146}$$

Now let us consider the individual components. For example, we shall find T_{11}:

$$\begin{aligned}T_{11}&=-\frac{1}{c}F_{1m}f_{1m}+\Lambda=-\frac{1}{c}F_{11}f_{11}-\frac{1}{c}F_{12}f_{12}-\frac{1}{c}F_{13}f_{13}-\\&-\frac{1}{c}F_{14}f_{14}+\Lambda=-\frac{1}{c}(cB_zH_z+cB_yH_y-cE_xD_x)+\frac{\boldsymbol{BH}}{2}-\frac{\boldsymbol{DE}}{2}=\\&=-\boldsymbol{BH}+H_xB_x+E_xD_x+\frac{\boldsymbol{BH}}{2}-\frac{\boldsymbol{DE}}{2}=\\&=H_xB_x+E_xD_x-\frac{\boldsymbol{DE}+\boldsymbol{BH}}{2}=H_1B_1+E_1D_1-w.\end{aligned} \tag{6.147}$$

We have found T_{11} to be a component of a three-dimensional tension tensor of Maxwell (6.128). Similarly, it can be shown that all components of the tensor $T_{\alpha\beta}$, that is the components whose indices i, k take on the values from 1 to 3, coincide with a tension tensor of Maxwell (6.128). It remains to consider the T_{ik} components in which at least one of the indices is equal to 4. We shall start from T_{44}:

$$\begin{aligned}T_{44}&=-\frac{1}{c}F_{4m}f_{4m}+\Lambda=\\&=E_xD_x+E_yD_y+E_zD_z+\frac{\boldsymbol{BH}}{2}-\frac{\boldsymbol{DE}}{2}=\frac{\boldsymbol{BH}}{2}+\frac{\boldsymbol{DE}}{2}=w.\end{aligned} \tag{6.148}$$

The component T_{44} turned out to be equal to the electromagnetic field energy density. Now let us find T_{14}:

$$\begin{aligned}T_{14}=T_{41}&=-\frac{1}{c}F_{1m}f_{4m}=-\frac{1}{c}(F_{12}f_{42}+F_{13}f_{43})=\\&=-ic(D_yB_z-D_zB_y)=-ic\varepsilon_0\mu_0[\boldsymbol{EH}]_x=\\&=-\frac{i}{c}S_x=-ic\frac{S_x}{c^2}=-icg_x.\end{aligned} \tag{6.149}$$

In much the same way

$$\begin{aligned}T_{24}&=T_{42}=-icg_y,\\T_{34}&=T_{43}=-icg_z.\end{aligned} \tag{6.150}$$

The components T_{14}, T_{24}, T_{34} turned out to be proportional to the components of the electromagnetic field momentum density $\boldsymbol{g}=\boldsymbol{S}/c^2$. Later on it will be clear (see Eq. (6.153)) that we deal here with a momentum density indeed, and not an energy flux to

which a momentum is proportional. Write out the matrix of the energy-momentum-tension tensor for a case of an electromagnetic field *in vacuo*:

$$T=\begin{vmatrix} T_{11} & T_{12} & T_{13} & -icg_x \\ T_{21} & T_{22} & T_{23} & -icg_y \\ T_{31} & T_{32} & T_{33} & -icg_z \\ -\frac{i}{c}S_x & -\frac{i}{c}S_y & -\frac{i}{c}S_z & w \end{vmatrix}=\begin{pmatrix} T_{\alpha\beta} & -icg \\ -\frac{i}{c}S & w \end{pmatrix}. \tag{6.151}$$

The upper left square comprising nine quantities defines a tension tensor of Maxwell. It becomes a correct quantity in a relativistic case when bordered with energy quantities S and w. Let us make sure that having composed the tensor T_{ik}, we obtained the energy and momentum conservation laws expressed in a three-dimensional form by Eqs. (6.125) and (6.132). Consider now the spatial components of a 4-force:

$$f_\alpha=\frac{\partial T_{\alpha\beta}}{\partial x_\beta}-\frac{i}{c}\frac{\partial S_\alpha}{\partial (ict)}=\frac{\partial T_{\alpha\beta}}{\partial x_\beta}-\frac{1}{c^2}\frac{\partial S_\alpha}{\partial t}=\frac{\partial T_{\alpha\beta}}{\partial x_\beta}-\frac{\partial g_\alpha}{\partial t}. \tag{6.152}$$

We took account of a three-dimensional momentum of an electromagnetic field *in vacuo* having the components $g_\alpha = S_\alpha/c^2$. Multiplying each component f_α $(\alpha = 1, 2, 3)$ by its respective unit vector m_α $(\alpha = 1, 2, 3)$ and summing up the values thus obtained, we get ($f = f_\alpha m_\alpha$ is a three-dimensional Lorentz force)

$$f+\frac{\partial g}{\partial t}=\frac{\partial T_{\alpha\beta}}{\partial x_\beta}m_\alpha. \tag{6.153}$$

Integrating the identity (6.153) over an arbitrary volume, we get

$$\int_{\mathcal{V}} f\,dV+\int_{\mathcal{V}}\frac{\partial g}{\partial t}\,d\mathcal{V}=\int_{\mathcal{V}}\frac{\partial T_{\alpha\beta}}{\partial x_\beta}m_\alpha\,d\mathcal{V}. \tag{6.154}$$

The left-hand side of Eq. (6 154) features a variation of a total momentum of particles as well as that of a total momentum of a field:

$$\int f\,d\mathcal{V}=\frac{dP}{dt},\quad \int\frac{\partial g}{\partial t}\,d\mathcal{V}=\frac{\partial}{\partial t}\int g\,d\mathcal{V}=\frac{\partial G}{\partial t}. \tag{6.155}$$

Applying the Gauss-Ostrogradsky theorem to the right-hand side of Eq. (6.154), we get

$$\int_{\mathcal{V}}\frac{\partial T_{\alpha\beta}}{\partial x_\beta}m_\alpha\,d\mathcal{V}=\oint_S T_{\alpha\beta}n_\beta m_\alpha\,dS=\oint_S T_{\alpha\beta}n_\alpha m_\beta\,dS; \tag{6.156}$$

the last transition makes use of the symmetry of the tensor $T_{\alpha\beta}$. Now then, we have arrived at the momentum conservation law (Eq. (6.132)) and have made sure that the components T_{14}, T_{24}, T_{34} are indeed proportional to electromagnetic field momentum components. The expression $T_{\alpha\beta}n_\alpha m_\beta$ can be considered not only as a force acting on a surface element, but also as a momentum flux through that surface element. The quantity $T_{\alpha\beta}m_\beta$ yields a vector component of this flux. Surely, both these interpretations are equivalent.

Consider now f_4. On the one hand, according to Eq. (6.54)

$$f_4 = \frac{i\rho}{c}(\boldsymbol{v}\boldsymbol{E}) = \frac{i}{c}(\boldsymbol{f}\boldsymbol{v}), \tag{6.157}$$

and on the other hand

$$f_4 = \frac{\partial T_{4k}}{\partial x_k} = \frac{\partial T_{41}}{\partial x_1} + \frac{\partial T_{42}}{\partial x_2} + \frac{\partial T_{43}}{\partial x_3} + \frac{\partial T_{44}}{\partial x_4} = -\frac{i}{c}\operatorname{div}\boldsymbol{S} + \frac{\partial w}{\partial(ict)}. \tag{6.158}$$

Consequently, Eq. (6.158) can be rewritten as

$$\frac{\partial w}{\partial t} + \operatorname{div}\boldsymbol{S} + (\boldsymbol{v}\boldsymbol{f}) = 0. \tag{6.159}$$

Integrating Eq. (6.159) with respect to an arbitrary volume of a field and taking account of Eqs. (6.121) and (6.123), we get

$$\frac{d}{dt}(T + W) = -\oint_S \boldsymbol{S}\, d\boldsymbol{S}, \tag{6.160}$$

with the Gauss theorem being applied to the term $\operatorname{div}\boldsymbol{S}$ of Eq. (6.159). This is precisely the energy conservation law (Eq. (6.125)).

Thus, in the relativistic theory the Maxwellian tensions, momentum and energy of a field *in vacuo* amalgamate into a single tensor quantity, the energy-momentum-tension tensor. The energy and momentum conservation laws manifest themselves via a single relation.

The symmetry of the energy-momentum-tension tensor constitutes a fundamentally important property. Owing to this, the fundamental relationship between the energy and momentum flux densities follows immediately for the case of an electromagnetic field *in vacuo*:

$$\boldsymbol{S} = \boldsymbol{g}c^2. \tag{6.161}$$

One can readily make sure that the spur of the tensor T_{ik}, i.e. the sum of its diagonal components, is equal to zero.

Having established a tensor character of tensions, of a momentum, and of an electromagnetic field energy flux and density, we

automatically obtain the rules according to which these quantities are transformed on transition from one inertial frame of reference to another. We shall write out only those transformation formulae that will be needed later. Substituting the component values from Eq. (6.151) into the general equations (A.I. 31), we get

$$T_{xx}=T_{11}=\Gamma^2\left(T'_{xx}-2\frac{V}{c^2}S'_x-\frac{V^2}{c^2}w'\right), \tag{6.162}$$

$$T_{44}=\Gamma^2\left(w'+2\frac{V}{c^2}S'_x-\frac{V^2}{c^2}T'_{xx}\right), \tag{6.163}$$

$$T_{xy}=T_{12}=\Gamma\left(T'_{xy}-\frac{V}{c^2}S'_y\right), \tag{6.164}$$

$$g_x=\Gamma^2\left\{\left(1+\frac{V^2}{c^2}\right)g'_x+\frac{V}{c^2}w'-\frac{V}{c^2}T'_{xx}\right\}, \tag{6.165}$$

$$g_y=\Gamma\left(g'_y-\frac{V}{c^2}T'_{xy}\right), \tag{6.166}$$

$$g_z=\Gamma\left(g'_z-\frac{V}{c^2}T'_{zx}\right). \tag{6.167}$$

§ 6.12. An energy-momentum-tension tensor of an electromagnetic field in a medium. The Minkowski tensor and Abraham tensor. An energy-momentum-tension tensor (EMT) in a medium attracts interest primarily because this tensor is associated with a momentum of an electromagnetic field in a medium. The latter quantity is directly related to the quantities observed in an experiment, e. g. to the pressure of light. This tensor, however, is not defined in a unique manner in a medium, and the discussion about its "proper form" is still going on.

Let us find the general form of an energy-momentum tensor in a uniform isotropic medium. It was shown in § 6.11 that the general form of energy-momentum tensor components in a uniform isotropic medium does not differ from that in the case of vacuum; in both cases (see Eq. (6.146))

$$T_{ik}=-\frac{1}{c}F_{im}f_{km}+\delta_{ik}\Lambda. \tag{6.168}$$

However, the proportionality factors between spatial and temporal components $\mathfrak{f}$ and $\mathfrak{F}$ are different in a medium (see Eqs. (6.140) and (6.141)), and the tensor T_{ik} defined according to Eq. (6.168) turns out to be asymmetric in contrast to the tensor (6.149). Asymmetry arises due to temporal components of the tensor; spatial components are symmetric, at least in an isotropic medium. It is indeed easy to see that spatial components of the tensor T_{ik} in a medium differ from those *in vacuo* only by the values of a

and μ. So, for example

$$T_{11} = -\frac{1}{c} F_{1m} f_{1m} + \Lambda =$$
$$= -\frac{1}{c}\left(cB_z \frac{B_z}{\mu} + cB_y \frac{B_y}{\mu} - c\varepsilon E_x^2\right) + \Lambda =$$
$$= -H_z B_z - H_y B_y - H_x B_x + H_x B_x + E_x D_x + \frac{BH - DE}{2} =$$
$$= H_x B_x + E_x D_x - \frac{ED + BH}{2}.$$

This expression coincides with the value of T_{11} given by Eq. (6.147), the only difference being the quantities ε and μ replacing the quantities ε_0 and μ_0 of Eq. (6.147). Exactly in the same way we shall find that

$$T_{44} = -\frac{1}{c} F_{4m} f_{4m} + \Lambda = \frac{ED + BH}{2} = w.$$

However, if

$$T_{14} = -ic\varepsilon\mu\,[\boldsymbol{EH}]_x = -(i/c)(\varepsilon\mu/\varepsilon_0\mu_0)\, S_x,$$

then

$$T_{41} = -(i/c)\,[\boldsymbol{EH}]_x = -(i/c)\, S_x.$$

Consequently, the energy-momentum-tension tensor in a uniform isotropic medium, obtained by a direct transformation of a 4-force, is not symmetric any more. It is called the Minkowski tensor, and its components have the form

$$T_{ik}^{\mathrm{M}} = \begin{pmatrix} T_{11} & T_{12} & T_{13} & -(i/c)\, n^2 S_x \\ T_{21} & T_{22} & T_{23} & -(i/c)\, n^2 S_y \\ T_{31} & T_{32} & T_{33} & -(i/c)\, n^2 S_z \\ -(i/c)\, S_x & -(i/c)\, S_y & -(i/c)\, S_z & w \end{pmatrix} =$$
$$= \begin{pmatrix} T_{\alpha\beta} & -(i/c)\, n^2 \boldsymbol{S} \\ -(i/c)\, \boldsymbol{S} & w \end{pmatrix}. \qquad (6.169)$$

Eq. (6.169) includes a refraction index $n = \sqrt{\varepsilon\mu/\varepsilon_0\mu_0}$. The momentum density corresponding to tensor (6.169) (i.e. the components T_{14}, T_{24}, T_{34}) turns out to be equal to

$$\boldsymbol{g}^{\mathrm{M}} = \frac{n^2}{c^2}\boldsymbol{S} = [\boldsymbol{DB}]. \qquad (6.170)$$

The superscript "M" appearing in a momentum density symbol points to the fact that this density corresponds to the Minkowski tensor. The momentum density of a field in a medium (see Eq. (6.170)) exceeds n^2 times that *in vacuo* (Eq. (6.133)).

It is often suggested to use Eq. (6.133) for the description of the momentum density of a field in a medium as well. Surely, this way we subdivide a total momentum into a momentum of a field and a momentum of a medium itself. However, a momentum density of a field can be separated in the form of Eq. (6.133) only if another energy-momentum tensor, different from the Minkowski tensor, is utilized. It is also important to keep Eq. (6.133) because it yields the most general formulation of the energy inertia law. Since an energy flux is described by the components $T_{4\alpha}$ of an energy-momentum tensor and a momentum density by the components $T_{\alpha 4}$, Eq. (6.133) indicates the tensor's symmetry. Therefore, we should construct a new symmetric tensor satisfying the following conditions: $T_{\alpha 4} = -icg_\alpha$ and $T_{4\alpha} = -(i/c)\, S_\alpha$; Eq. (6.133) should be satisfied as well. The three-dimensional tension tensor $T_{\alpha\beta}$ should coincide with the three-dimensional tension tensor of Maxwell (6.128). Such a tensor was suggested by Abraham in the following form:

$$T^{\mathrm{A}}_{ik} = \begin{pmatrix} T_{\alpha\beta} & -icg^{\mathrm{A}} \\ -(i/c)\,S & w \end{pmatrix}, \quad g^{\mathrm{A}} = g = \frac{S}{c^2}. \tag{6.171}$$

Since $T_{\alpha\beta} = T_{\beta\alpha}$ and owing to Eq. (6.133), this tensor is symmetric. The introduction of the Abraham tensor implies an appearance of a volume force acting on a medium. This force is referred to as the Abraham force. To find its magnitude, recall that the Lorentz force density components are related to the Minkowski energy-momentum tensor by the ratio

$$f_\alpha = \frac{\partial T^{\mathrm{M}}_{\alpha k}}{\partial x_k}. \tag{6.172}$$

This is precisely how the Minkowski tensor was obtained.

It follows directly from Eq. (6.172) that

$$f_\alpha m_\alpha \equiv f^{\mathrm{L}} = \frac{\partial T^{\mathrm{M}}_{\alpha k}}{\partial x_k} m_\alpha. \tag{6.173}$$

Writing out the sum on the right-hand side and interchanging the terms of the equation, we get (see Eq. (6.169))

$$\frac{\partial T^{\mathrm{M}}_{\alpha\beta}}{\partial x_\beta} m_\alpha - \frac{\partial g^{\mathrm{M}}}{\partial t} = f^{\mathrm{L}}, \tag{6.174}$$

where f^{L} denotes the Lorentz force density.

Now let us write the expression $(\partial T^{\mathrm{A}}_{\alpha k}/\partial x_k)\, m_\alpha$ in full, taking into account Eq. (6.171):

$$\frac{\partial T^{\mathrm{A}}_{\alpha k}}{\partial x_k} m_\alpha = \frac{\partial T_{\alpha\beta}}{\partial x_\beta} m_\alpha - \frac{\partial g^{\mathrm{A}}}{\partial t} \equiv \frac{\partial T_{\alpha\beta}}{\partial x_\beta} m_\alpha - \frac{\partial g^{\mathrm{M}}}{\partial t} + \frac{\partial}{\partial t}(g^{\mathrm{M}} - g^{\mathrm{A}}). \tag{6.175}$$

The second link of Eq. (6.175) takes into account that the tension tensor of Maxwell is the same both in the Minkowski and in the Abraham tensor; the third link of Eq. (6.175) is the identical transcription of the second link. But the first two terms of the last link of Eq. (6.175) can now be substituted according to Eq. (6.174) to yield

$$\frac{\partial T_{\alpha\beta}}{\partial x_\beta} m_\alpha - \frac{\partial g^A}{\partial t} = f^L + f^A. \tag{6.176}$$

The right-hand side of Eq. (6.176) contains a term f^A representing the derivative of a momentum density; according to the second law of Newton the derivative of a momentum density with respect to time makes up a force density

$$f^A = \frac{\partial}{\partial t}(g^M - g^A) = \frac{\partial}{\partial t}\left\{[\boldsymbol{DB}] - \frac{1}{c^2}[\boldsymbol{EH}]\right\}. \tag{6.177}$$

In an isotropic medium

$$f^A = \frac{\partial}{\partial t}(\varepsilon\mu - \varepsilon_0\mu_0)\,\boldsymbol{S} = \frac{\partial}{\partial t}\left(\frac{n^2-1}{c^2}\,\boldsymbol{S}\right). \tag{6.178}$$

A force density given by Eqs. (6.177) and (6.178) is referred to as the Abraham force density.

The Minkowski tensor and the Abraham tensor furnish different expressions for an electromagnetic field momentum density. Let us write out the corresponding expressions for a plane electromagnetic wave momentum density. In the case of a plane electromagnetic wave propagating in a uniform isotropic dielectric the relationship between the value of the Poynting vector S, the monochromatic wave phase velocity v and the wave energy density w is given by the simple equation:

$$S = w \cdot v, \tag{6.179}$$

$$v = c/n = (\varepsilon\mu/\varepsilon_0\mu_0)^{1/2}. \tag{6.180}$$

Eqs. (6.170) and (6.171) furnish the momentum densities:

$$g^M = (n^2/c^2)\,S = (n^2/c^2)\,w\,(c/n) = (w/c)\,n, \tag{6.181}$$

$$g^A = (S/c^2) = (1/c^2)w\,(c/n) = (w/cn). \tag{6.182}$$

Let us assume that electromagnetic field energy can be quantized, i.e. $w = N\hbar\omega$, where N is the number of quanta in a unit of volume. Then from Eqs. (6.181) and (6.182) we obtain the following quantum momenta in a medium:

$$p^M = (\hbar\omega/c)\,n, \tag{6.183}$$

$$p^A = (\hbar\omega/cn). \tag{6.184}$$

Which of these two equations is "true"? The technique of secondary quantization of an electromagnetic field in matter results in Eq. (6.183). Let us assume that a quantum momentum in a medium $\boldsymbol{p}$ is defined as follows:

$$\boldsymbol{p} = (\hbar\omega/c)\, n\boldsymbol{s}, \tag{6.185}$$

where $\boldsymbol{s}$ is a unit vector in the direction of wave propagation, and a 4-vector of a quantum energy-momentum takes the form

$$\vec{p}\left(\frac{\hbar\omega}{c} n\boldsymbol{s},\ \frac{\hbar\omega}{c}\right). \tag{6.186}$$

Using Eq. (6.186), one can obtain a correct expression stipulating the Vavilov-Cherenkov radiation (see Chapter 7). This circumstance seems to indicate that Eq. (6.186) and the Minkowski tensor are to be preferred. However, a meticulous usage of any of these tensors provides a correct result. The point is that both tensors satisfy Eq. (6.127) which is the consequence of the field equations. It is important to know how an electromagnetic field momentum is defined in matter. In a general case, it is impossible to split the total momentum of a field into a fraction pertaining only to matter and that pertaining only to a field. But this is precisely what it is done when the Abraham momentum is introduced. When a light wave passes from vacuum into a medium, its momentum is not totally transferred into this medium; a fraction of the momentum is transmitted to the medium itself. The Minkowski tensor is utilized whenever a total transmitted momentum is considered; if we deal with a momentum related to radiation in a medium, the Abraham tensor should be used. Eq. (6.185) provides a correct value for a photon momentum in the case of the Vavilov-Cherenkov radiation (see Chapter 7) because an overall momentum which a Cherenkov electron transmits to a medium is of prime importance here. The overall momentum transmitted to a photon in a medium is just equal to $\hbar\omega n/c$. No wonder that quantization of an electromagnetic field in dielectrics results in Eq. (6.185) describing a momentum. This expression represents an overall momentum of an electromagnetic field, this momentum being associated both with a field and with a matter (see § 7.7).

As to a force acting on a matter, it is related to the Abraham force and, naturally, to the Abraham tensor. An attempt to measure the Abraham force was undertaken in 1975 which apparently succeeded.

Fig. 6.7 illustrates the experimental arrangement. A disc made of barium titanate ($\varepsilon \approx 4000$, $\mu \approx \mu_0$) has a small hole in the centre. Both rims of the disc are coated with aluminium; consequently, it forms a cylindrical capacitor. The disc is so suspended

on a long tungsten filament that it can perform torsional oscillations between the poles of a dc electromagnet generating a 10 kGs field. An ac voltage with a 150 V peak value is applied to the internal rim electrode, while the outer electrode is grounded by means of a thin gold wire not affecting the disc's oscillations. The voltage is applied in phase with the characteristic oscillations of the disc.

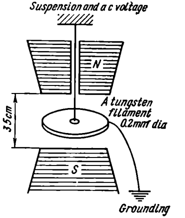


Fig. 6.7. An experimental observation of the Abraham force.

The Abraham force can also be written in the form

$$f^A = \varepsilon_0\mu_0(\varkappa_m\varkappa_e - 1)\frac{\partial S}{\partial t} =$$
$$= \varepsilon_0\mu_0(\varkappa_m\varkappa_e - 1)[\dot{E}H], \quad (6.187)$$

where

$$\varkappa_e = \varepsilon/\varepsilon_0, \quad \varkappa_m = \mu/\mu_0 \quad (6.188)$$

and the fact that the magnetic field is constant ($\dot{H} = 0$) is taken into accouht. In the case of barium titanate $\varkappa_m \approx 1$ and consequently

$$f^A = [\varepsilon_0(\varkappa_e - 1)\dot{E}, \mu_0 H] =$$
$$= [\dot{P}, \mu_0 H]; \quad (6.189)$$

the last equation follows from the fact that in a uniform isotropic medium

$$P = D - \varepsilon_0 E = (\varepsilon - \varepsilon_0)E = \varepsilon_0(\varkappa_e - 1)E. \quad (6.190)$$

The physical meaning of the "Abraham force" is obvious in the specific case just considered. $\dot{P}$ is the fraction of the displacement current caused by the motion of bound charges. In essence, this is the Ampere force. It is not difficult to find out that this force induces a torque, the electric field being directed radially. Surely, the essential point is that no other forces associated with the presence of an electromagnetic field contribute to the torque.

The authors who proposed this experiment claim that the observed oscillations of the disc are consistent with the calculation based on the presence of the Abraham force. We should point out here again that this experimental result, however interesting *per se*, by no means "chooses" between the tensors (6.169) and (6.171). Some additional remarks concerning a choice of an expression for a photon momentum in a medium can be found in § 7.7.

§ 6.13. An energy-momentum-tension tensor of a spherically symmetric charge. If an electric charge *in vacuo* is at rest in the frame K, there is only an electric field in the frame and the energy-

momentum-tension tensor can readily be written in the form

$$T_{ik}=\begin{pmatrix} T_{\alpha\beta} & 0 \\ 0 & w \end{pmatrix}. \tag{6.191}$$

where $w=\varepsilon_0 E^2/2$ and $T_{\alpha\beta}=\varepsilon_0 E_\alpha E_\beta-\delta_{\alpha\beta}w$. If the charge moves relative to the frame K' at the velocity $-V$, its tensor T'_{ik} can be found through the use of the general equations for the tensor component transformation. Specifically, in order to transform the momentum density along the x axis, that is $(i/c)\ T_{14}$, and the energy density T_{44}, we have the following equations (see A.I. 31):

$$T'_{14}=-iB\Gamma^2 T_{11}+\Gamma^2 iBT_{44}=iB\Gamma^2(w-T_{11}),$$

$$T'_{44}=\Gamma^2 T_{44}-B^2\Gamma^2 T_{11}=\Gamma^2\left(w-\frac{V^2}{c^2}T_{11}\right).$$

Let us find the total energy and total momentum of a point charge, taking into account that a transition from a volume element $d\mathscr{V}'$ in the frame K' to a volume element $d\mathscr{V}$ in the frame K is effected by means of the formula $d\mathscr{V}'=\frac{1}{\Gamma}d\mathscr{V}$:

$$U'=\int T'_{44}\,d\mathscr{V}'=\int T'_{44}\frac{1}{\Gamma}\,d\mathscr{V}=\Gamma\int(w-B^2T_{11})\,d\mathscr{V}, \tag{6.192}$$

$$G'_x=\frac{i}{c}\int T'_{14}\,d\mathscr{V}'=\frac{i}{c}\int T'_{14}\frac{1}{\Gamma}\,d\mathscr{V}=\frac{B\Gamma}{c}\int(T_{11}-w)\,d\mathscr{V}. \tag{6.193}$$

It is obvious that

$$\int w\,d\mathscr{V}=\int\frac{\varepsilon_0 E^2}{2}\,d\mathscr{V}=U.$$

If in the charge's proper frame of reference K it possesses a spherical symmetry, so that $E_x^2=E_y^2=E_z^2=E^2/3$, then

$$\int T_{11}\,d\mathscr{V}=\varepsilon_0\int\left(E_x^2-\frac{E^2}{2}\right)d\mathscr{V}=-\frac{\varepsilon_0}{6}\int E^2\,d\mathscr{V}=-\frac{U}{3}.$$

Consequently, Eqs. (6.192) and (6.193) take the form

$$U'=\Gamma U\left(1+\frac{B^2}{3}\right), \tag{6.194}$$

$$G'_x=-\frac{4}{3}\frac{B}{c}\Gamma U. \tag{6.195}$$

The momentum components G'_y and G'_z turn into zero, so that

$$\boldsymbol{G}'=-\frac{4}{3}\frac{\boldsymbol{V}}{c^2}\Gamma U. \tag{6.196}$$

The appearance of the "minus" sign in the last equation is due to the fact that the charge moves at the velocity $-V$ relative to the

frame K'. Comparing Eqs. (6.194) and (6.195) with the equations transforming a momentum and energy of a particle on transition from the proper frame of reference to an arbitrary one (see Eq. (5.49)), we see that these equations are different. In former times an effort was made to treat an electron mass as an electromagnetic one using the relation

$$m = U/c^2. \tag{6.197}$$

Eq. (6.196) shows that such an interpretation gets into trouble since to "confine" a charge some additional forces are required neutralizing repulsion, i.e. the additional energy that was not accounted for. Having taken into account mechanical stresses, we can obtain the following relations:

$$\boldsymbol{G}' = -\Gamma \boldsymbol{V} \frac{U}{c^2}, \quad U' = \Gamma U \tag{6.198}$$

in a complete agreement with Eq. (5.49). The details can be seen in [13].

§ 6.14. The field potentials in a moving non-conducting medium.* In § 6.1 we introduced a 4-potential of an electromagnetic field *in vacuo*. Surely, an electromagnetic field can be determined immediately from the Maxwell equations, without resorting to potentials. In many cases, however, the utilization of potentials as intermediate quantities determining the fields $\boldsymbol{E}$ and $\boldsymbol{B}$, proves to be very convenient if only because the number of functions to be determined decreases. Knowing only four components of a vector potential, one can find from them all components of an electric and a magnetic field. The potentials are still more convenient to use in electrodynamics of moving media where the material equations (6.74) and (6.75) turn out to be much more complicated than in the case of a stationary medium.

It will be shown below how one can obtain the expressions for field potentials in a moving medium. To illustrate an application of such potentials, we shall consider the propagation of a plane electromagnetic wave in a medium moving relative to a stationary observer. This example has a direct bearing on problems analysed in Chapter 7. In this section we apply the methods of tensor algebra described briefly in Appendix I, § 3.

Now let us derive equations for a 4-potential in moving media. We shall be describing a field in a moving medium by two tensors: F_{ik} (see Eq. (6.29)) and f_{ik} (see Eq. (6.31)). The tensor F_{ik} is sometimes referred to as a *field tensor*, and the tensor f_{ik} as an *induction tensor*.

---

* §§ 6.14 and 6.15 are written by B. M. Bolotovsky and S. N. Stolyarov.

Let us introduce a four-dimensional field potential in a medium, Φ, having defined it by the following relation

$$F_{ik}=c\left(\frac{\partial\Phi_k}{\partial x_i}-\frac{\partial\Phi_i}{\partial x_k}\right), \tag{6.199}$$

coinciding with Eq. (6.28). Knowing the four components of the potential Φ_k, we can use this equation to determine all components of the tensor F_{ik}, that is the magnetic induction $\boldsymbol{B}$ and the electric field $\boldsymbol{E}$. In order to describe wholly an electromagnetic field in a medium, one needs to know also the components of the tensor f_{ik}, i.e. the components of the vectors of the magnetic field and electric induction. If the field tensor $\mathfrak{F}$ is known, the induction tensor $\mathfrak{f}$ can be determined by means of the material equations (6.79) and (6.80) relating the components of these two tensors. Recall that the correlation between the tensors $\mathfrak{F}$ and $\mathfrak{f}$ in a vector form is given by the Minkowski equations (6.74) and (6.75).

The material equations (6.79) and (6.80) defining the relationship between the tensors $\mathfrak{F}$ and $\mathfrak{f}$ can be written in the form of a single tensor relation

$$f_{ik}=\varepsilon_{iklm}F_{lm}, \tag{6.200}$$

where the tensor of the fourth rank ε_{iklm} is so chosen as to make the Minkowski equations (6.74) and (6.75) valid. It is easy to show that the tensor of the form

$$\varepsilon_{iklm}=\frac{1}{\mu c}\left(\delta_{il}-\varkappa c^{-2}U_iU_l\right)\left(\delta_{km}-\varkappa c^{-2}U_kU_m\right) \tag{6.201}$$

possesses the necessary properties. Here δ_{il} is the Kronecker delta defined by Eq. (A.I. 4), and U_k are the four-dimensional velocity components. This four-dimensional velocity was discussed in Chapter 5 and shown to have the components $\vec{V}(\Gamma\boldsymbol{V}, ic\Gamma)$ where $\boldsymbol{V}$ is the three-dimensional velocity of a medium motion. The dimensionless constant $\varkappa$ is defined via the refraction index n:

$$\varkappa=\frac{\varepsilon\mu}{\varepsilon_0\mu_0}-1=n^2-1, \quad n=\frac{c}{v}=\sqrt{\frac{\varepsilon\mu}{\varepsilon_0\mu_0}}. \tag{6.202}$$

It is easy to see that $\varkappa=0$ *in vacuo*, and the relation (6.200) between the tensors $\mathfrak{f}$ and $\mathfrak{F}$ takes the form

$$f_{ik}=\frac{1}{\mu_0 c}F_{ik}, \tag{6.203}$$

corresponding to the well-known relations between the fields $\boldsymbol{E}$ and $\boldsymbol{H}$, and the inductions $\boldsymbol{D}$ and $\boldsymbol{B}$ *in vacuo*:

$$\boldsymbol{D}=\varepsilon_0\boldsymbol{E}, \quad \boldsymbol{B}=\mu_0\boldsymbol{H}. \tag{6.204}$$

In a stationary medium the tensor ε_{iklm} of Eq. (6.201) yields the following relations between the fields and inductions:

$$\boldsymbol{D} = \varepsilon \boldsymbol{E}, \quad \boldsymbol{B} = \mu \boldsymbol{H}. \tag{6.205}$$

This is easy to ascertain, having assumed $U_1 = U_2 = U_3 = 0$ and $U_4 = ic$ in Eq. (6.201).

Since the components of the tensor F_{lm} are expressed via the components of the four-dimensional potential Φ_l, and the components of the induction tensor f_{ik} are related to F_{lm} by Eq. (6.200), the components of the tensor f_{ik} can also be expressed through the components of the four-dimensional potential Φ_l. In other words, to determine all components of fields and inductions in moving media, it proves to be sufficient to know the four functions Φ_l.

Now let us derive the equations for field potentials in moving media. For this purpose let us make use of Eq. (6.60):

$$\frac{\partial f_{ik}}{\partial x_k} = s_i.$$

Substituting f_{ik} from Eq. (6.200), we get

$$\varepsilon_{iklm} \frac{\partial F_{lm}}{\partial x_k} = s_i. \tag{6.206}$$

Utilizing Eq. (6.201) which gives an explicit expression for the tensor ε_{iklm}, as well as Eq. (6.28) expressing F_{lm} through the potential components Φ_l, we reduce Eq. (6.206) to the following form by simple transformations:

$$\frac{1}{\mu c}(\delta_{il} - \varkappa c^{-2} U_i U_l) \left\{ \frac{\partial}{\partial x_k} \left[\frac{\partial \Phi_k}{\partial x_l} - \varkappa c^{-2} U_k U_m \frac{\partial \Phi_m}{\partial x_k} \right] - \right.$$
$$\left. - \left[\frac{\partial^2}{\partial x_k^2} - \varkappa c^{-2} \left(U_k \frac{\partial}{\partial x_k} \right)^2 \right] \Phi_l \right\} = s_i. \tag{6.207}$$

Now multiplying both sides of Eq. (6.207) by the tensor

$$\left(\delta_{i\alpha} + \frac{\varkappa}{1+\varkappa} c^{-2} U_i U_\alpha \right)$$

and making use of the relationship

$$(\delta_{il} - \varkappa c^{-2} U_i U_l) \left(\delta_{i\alpha} + \frac{\varkappa}{1+\varkappa} c^{-2} U_i U_\alpha \right) = \delta_{l\alpha}, \tag{6.208}$$

which is easy to check, we finally obtain

$$\left[\frac{\partial^2}{\partial x_k^2} - \varkappa c^{-2} \left(U_k \frac{\partial}{\partial x_k} \right)^2 \right] \Phi_\alpha - \frac{\partial}{\partial x_\alpha} \left(\frac{\partial \Phi_k}{\partial x_k} - \varkappa c^{-2} U_k U_m \frac{\partial \Phi_m}{\partial x_k} \right) =$$
$$= -\mu c \left(\delta_{i\alpha} + \frac{\varkappa}{1+\varkappa} c^{-2} U_i U_\alpha \right) s_i. \tag{6.209}$$

The system of equations (6.209) defines all components of the potential Φ_α from the given field sources s_l in a moving medium.

This system can be simplified provided a well chosen additional condition is imposed on the potentials, e.g. it is required that the following relation is to be valid:

$$\frac{\partial \Phi_k}{\partial x_k} - \varkappa c^{-2} U_k U_m \frac{\partial \Phi_m}{\partial x_k} = 0. \tag{6.210}$$

This condition is a generalization of the well-known Lorentz condition imposed on potentials *in vacuo* (see Eq. (6.8)). The feasibility of Eq. (6.210) is demonstrated in the same way as in conventional electrodynamics.

When condition (6.210) is satisfied, Eq. (6.209) becomes simplified and takes the form

$$\left\{ \frac{\partial^2}{\partial x_k^2} - \varkappa c^{-2} \left(U_k \frac{\partial}{\partial x_k} \right)^2 \right\} \Phi_\alpha = - \mu c \left(\delta_{l\alpha} + \frac{\varkappa}{1+\varkappa} c^{-2} U_l U_\alpha \right) s_l. \tag{6.211}$$

The system (6.211) is more convenient as compared to the system (6.209) since it comprises four equations, each of which includes only one component of a vector potential ($\alpha = 1, 2, 3, 4$). For given external sources, the solution of the system (6.211) fully defines the field generated by these sources in a moving medium.

If a moving medium has an interface, the system (6.211) should be supplemented by requisite boundary conditions (see § 6.8).

As an example of solving the equations obtained, let us consider an electromagnetic field in a moving medium in the absence of external sources, both currents and charges. Since in this case all $s_l = 0$, the system (6.211) turns into a system of the four uniform equations:

$$\left\{ \frac{\partial^2}{\partial x_k^2} - \varkappa c^{-2} \left(U_k \frac{\partial}{\partial x_k} \right)^2 \right\} \Phi_\alpha = 0. \tag{6.212}$$

Due to the additional condition (6.210) only three out of the four quantities Φ_α are independent. Accordingly, we can assume $\Phi_4 = 0$, and treat the remaining three quantities Φ_1, Φ_2, Φ_3 as components of some tensor which we shall denote by $\boldsymbol{A}$. Thus we see that in the case of such a calibration the vector potential Φ_α for a moving medium is a three-dimensional vector potential $\boldsymbol{A}$.

In this case the system of equations (6.212) can be rewritten in terms of the potential $\boldsymbol{A}$:

$$\left\{ \left(\Delta - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) - \frac{n^2 - 1}{c^2 (1 - \mathrm{B}^2)} \left[(\boldsymbol{V} \nabla) + \frac{\partial}{\partial t} \right]^2 \right\} \boldsymbol{A} = 0, \tag{6.213}$$

where $\mathrm{B} = V/c$, under the additional condition

$$\operatorname{div} \boldsymbol{A} - \frac{n^2 - 1}{c^2 (1 - \mathrm{B}^2)} \left[(\boldsymbol{V} \nabla) + \frac{\partial}{\partial t} \right] (\boldsymbol{V} \boldsymbol{A}) = 0, \tag{6.214}$$

which follows from the additional condition (6.210) at $\Phi_4 = 0$. If we know how to solve Eq. (6.213) with respect to the potential $\boldsymbol{A}$, the fields $\boldsymbol{E}$ and $\boldsymbol{B}$ can be expressed through $\boldsymbol{A}$ according to Eq. (6.28) which in our case takes a simple form

$$\boldsymbol{B} = \operatorname{rot} \boldsymbol{A}, \quad \boldsymbol{E} = -\frac{\partial \boldsymbol{A}}{\partial t}. \tag{6.215}$$

Knowing $\boldsymbol{E}$ and $\boldsymbol{B}$, we can find $\boldsymbol{D}$ and $\boldsymbol{H}$ by means of the Minkowski equations (6.74) and (6.75) for a moving medium.

Eq. (6.213) describes propagation of free electromagnetic waves in a moving medium. Free electromagnetic waves usually imply a field in the absence of charges and currents. Now let us pass over to the solution of this equation. We shall be seeking the solution for the vector potential $\boldsymbol{A}$ in the form of a plane electromagnetic wave:

$$\boldsymbol{A} = \boldsymbol{A}_0 e^{i(\omega t - \boldsymbol{kr})}. \tag{6.216}$$

Substituting this expression into Eq. (6.213), we obtain

$$\left\{ \left(-k^2 + \frac{\omega^2}{c^2} \right) + \frac{n^2 - 1}{c^2 (1 - \mathrm{B}^2)} (\boldsymbol{kV} - \omega)^2 \right\} \boldsymbol{A}_0 e^{i(\omega t - \boldsymbol{kr})} = 0. \tag{6.217}$$

Eq. (6.217) shows that the plane wave amplitude $\boldsymbol{A}_0$ differs from zero only for those waves that satisfy the condition

$$\left(-k^2 + \frac{\omega^2}{c^2} \right) + \frac{n^2 - 1}{c^2 (1 - \mathrm{B}^2)} (\boldsymbol{kV} - \omega)^2 = 0. \tag{6.218}$$

Eq. (6.218) can be readily derived from the dispersion equation which is valid for plane monochromatic waves in a stationary medium:

$$k^2 - \omega^2/v^2 = 0, \quad k^2 = \boldsymbol{k}^2.$$

We shall rewrite it in the form

$$\left(k^2 - \frac{\omega^2}{c^2} \right) - \frac{n^2 - 1}{c^2} \omega^2 = 0.$$

Here the parentheses enclose the square of the four-dimensional wave vector *in vacuo* $k\left(\boldsymbol{k}, i\frac{\omega}{c}\right)$, the quantity which is invariant relative to the Lorentz transformation. The quantity enclosed in the parentheses retains its appearance and numerical value in all inertial frames of reference. The second addendum of the last equation transforms as the frequency ω. In the reference frame in which a medium moves at the velocity $\boldsymbol{V}$, ω should be replaced by

$$\omega' = (\omega - \boldsymbol{kV})/\sqrt{1 - V^2/c^2}$$

(see § 7.2).

Due to these considerations the dispersion equation takes the form (6.218)

$$k^2 - \frac{\omega^2}{c^2} - \frac{n^2-1}{c^2(1-\mathrm{B}^2)}(\omega - \boldsymbol{kV})^2 = 0$$

in the reference frame relative to which the medium moves at the velocity $\boldsymbol{V}$. This condition defines a relationship between the wave vector $\boldsymbol{k}$ and frequency ω of a plane electromagnetic wave propagating in a moving medium. The additional condition (6.214) for such a wave takes the form

$$\left(\boldsymbol{A}_0,\ \boldsymbol{k} + \boldsymbol{V}\frac{n^2-1}{c^2(1-\mathrm{B}^2)}(\omega - \boldsymbol{kV})\right) = 0. \qquad (6.219)$$

From Eq. (6.219) requiring that the scalar product turn into zero, it follows that in a moving medium the vector $\boldsymbol{A}_0$ is perpendicular not to the wave propagation direction defined by the wave vector $\boldsymbol{k}$, but to a linear combination of the wave vector $\boldsymbol{k}$ and the velocity vector $\boldsymbol{V}$ of the medium. In the two specific cases, when a wave propagates *in vacuo* ($n = 1$) and when a medium is stationary ($\boldsymbol{V} = 0$), Eq. (6.219) turns into the well-known relation of free transverse electromagnetic waves: $\boldsymbol{A}_0\boldsymbol{k} = 0$ from which it follows that in a free electromagnetic wave the vectors $\boldsymbol{E}$, $\boldsymbol{H}$, $\boldsymbol{B}$ and $\boldsymbol{D}$ are perpendicular to the wave vector, that is to the wave propagation direction. In a moving medium, however, the waves are not transverse, generally speaking. Indeed, in the case of a plane wave (Eq. (6.216)) the fields $\boldsymbol{E}$ and $\boldsymbol{B}$ are determined according to Eq. (6.215):

$$\boldsymbol{B} = -i\,[\boldsymbol{kA}_0]\,e^{i(\omega t - \boldsymbol{kr})}, \quad \boldsymbol{E} = -i\omega\boldsymbol{A}_0 e^{i(\omega t - \boldsymbol{kr})}.$$

Whence it is seen that the vector $\boldsymbol{B}$ is perpendicular to the wave vector $\boldsymbol{k}$ while the vector $\boldsymbol{E}$ is not (since the vector $\boldsymbol{A}_0$ is not transverse according to the condition (6.219)).

Eq. (6.217) relating the wave vector $\boldsymbol{k}$ and frequency ω of a wave in a moving medium includes the scalar product $\boldsymbol{kV}$. This means that the wave propagation conditions depend on the angle between the propagation direction, or the wave vector $\boldsymbol{k}$, and the velocity of a medium $\boldsymbol{V}$. This circumstance indicates the phenomenon of carrying away of light by a moving medium. Let us consider this phenomenon in detail in the case of small velocities. Since the quantity $\mathrm{B} = V/c$ is small, we shall ignore all values of B having degrees higher than one in Eq. (6.128). We shall get

$$\frac{\omega^2}{c^2} - k^2 + \frac{n^2-1}{c^2}[\omega^2 - 2\omega(\boldsymbol{kV})] = 0,$$

or

$$\omega^2 - 2\boldsymbol{kV}\left(1 - \frac{1}{n^2}\right)\omega - \frac{k^2}{\varepsilon\mu} = 0.$$

Solving the obtained quadratic equation with respect to ω in the same approximation, we get

$$\omega = \pm \frac{k}{\sqrt{\varepsilon\mu}} + kV\left(1 - \frac{1}{n^2}\right). \tag{6.220}$$

From the two signs in front of the first addendum on the right-hand side one must choose the plus sign since in the case of $V = 0$ we have to get the well-known relationship between ω and k in a stationary medium.

$$\frac{\omega}{k} = \frac{1}{\sqrt{\varepsilon\mu}} = \frac{c}{n}; \tag{6.221}$$

here we introduced the refraction index of a stationary medium n. The quantity c/n is the phase velocity of light in a stationary medium.

The angle between the vectors k and V being denoted by θ, Eq. (6.220) takes the following form

$$\frac{\omega}{k} = \frac{c}{n} + V\cos\theta\left(1 - \frac{1}{n^2}\right) \tag{6.222}$$

for the indicated choice of a sign.

Just as in the case of a stationary medium (see Eq. (6.221)), the quantity ω/k in Eq. (6.222) defines the phase velocity of light but this time in a moving isotropic medium. Comparing Eqs. (6.222) and (6.221), we see that the phase velocity of light in a moving medium is different in different directions. If light propagates *along the motion* of a medium ($\cos\theta = 1$), the phase velocity is equal to

$$\frac{\omega}{k} = \frac{c}{n} + \left(1 - \frac{1}{n^2}\right)V.$$

If light propagates *against the motion* of a medium ($\cos\theta = -1$), then

$$\frac{\omega}{k} = \frac{c}{n} - \left(1 - \frac{1}{n^2}\right)V.$$

The factor $(1 - 1/n^2)$ is the so-called light *drag coefficient* which was experimentally measured by Fizeau with water serving as a moving medium.

§ 6.15. The field potentials in a moving conducting medium. Prior to dealing with field equations in a moving conducting medium we shall recall the main facts concerning the propagation of waves through a stationary medium in the presence of conductance.

In this case the Maxwell equations take the form

$$\begin{aligned} &\operatorname{rot} \boldsymbol{E} = -\dot{\boldsymbol{B}}, \quad \operatorname{rot} \boldsymbol{H} = \dot{\boldsymbol{D}} + \boldsymbol{j} + \sigma\boldsymbol{E}, \\ &\operatorname{div} \boldsymbol{D} = \rho, \quad \operatorname{div} \boldsymbol{B} = 0. \end{aligned} \tag{6.223}$$

Here ρ and j are the charge density and current density induced by "extraneous" sources. Subsequently we shall confine ourselves to consideration of equations (6.223) in the absence of "extraneous" sources, i.e. assume $\rho = 0$ and $j = 0$. The solution of equations (6.223) will be assumed in the form

$$\begin{aligned} \boldsymbol{E} &= \boldsymbol{E}_0 e^{i(\boldsymbol{kr}-\omega t)}, & \boldsymbol{D} &= \boldsymbol{D}_0 e^{i(\boldsymbol{kr}-\omega t)}, \\ \boldsymbol{H} &= \boldsymbol{H}_0 e^{i(\boldsymbol{kr}-\omega t)}, & \boldsymbol{B} &= \boldsymbol{B}_0 e^{i(\boldsymbol{kr}-\omega t)}, \end{aligned} \tag{6.224}$$

where $\boldsymbol{E}_0$, $\boldsymbol{D}_0$, $\boldsymbol{H}_0$, $\boldsymbol{B}_0$ are constant amplitudes that do not depend on either coordinates or time. Thus we seek the solution as a plane wave with the wave vector $\boldsymbol{k}$ and frequency ω.

Substituting the expressions for fields (6.224) into equations (6.223) and taking into account that $\rho = 0$ and $j = 0$, we obtain the following algebraic, not differential, equations interrelating amplitudes of fields:

$$\begin{aligned} [\boldsymbol{k}, \boldsymbol{E}_0] &= \omega \boldsymbol{B}_0, & [\boldsymbol{k}, \boldsymbol{H}_0] &= -\omega \boldsymbol{D}_0 - i\sigma \boldsymbol{E}_0, \\ (\boldsymbol{k}\boldsymbol{D}_0) &= 0, & (\boldsymbol{k}\boldsymbol{B}_0) &= 0. \end{aligned} \tag{6.225}$$

We have made use of the following equations:

$$\operatorname{rot}(\boldsymbol{E}_0 e^{i\boldsymbol{kr}}) = i[\boldsymbol{k}, \boldsymbol{E}_0] e^{i\boldsymbol{kr}},$$
$$\operatorname{div}(\boldsymbol{E}_0 e^{i\boldsymbol{kr}}) = i(\boldsymbol{k}\boldsymbol{E}_0) e^{i\boldsymbol{kr}}.$$

Eqs. (6.225) should be supplemented with material equations defining the relationship between field and induction values. In the simplest case of an isotropic stationary medium we assume the following relations:

$$\boldsymbol{D}_0 = \varepsilon \boldsymbol{E}_0, \quad \boldsymbol{B}_0 = \mu \boldsymbol{H}_0. \tag{6.226}$$

We shall suppose that the quantities ε and μ do not depend on field amplitudes. In so doing, we ensure a linear relationship between the fields $\boldsymbol{E}_0$, $\boldsymbol{H}_0$ and inductions $\boldsymbol{D}_0$, $\boldsymbol{B}_0$ respectively. However, ε and μ may, generally speaking, depend not only on field values but also on the frequency ω and wavelength $\lambda = 2\pi/k$. When ε and μ depend only on frequency, we observe a *frequency dispersion*. But if ε and μ depend on a wavelength, the medium is said to possess a *spatial dispersion*.

If the material equations (6.226) are substituted into Eqs. (6.225), one gets the equations involving only the field amplitudes $\boldsymbol{E}_0$ and $\boldsymbol{H}_0$:

$$\begin{aligned} [\boldsymbol{k}, \boldsymbol{E}_0] &= \mu\omega \boldsymbol{H}_0, & [\boldsymbol{k}\boldsymbol{H}_0] &= -\varepsilon\omega \boldsymbol{E}_0 - i\sigma \boldsymbol{E}_0, \\ \varepsilon(\boldsymbol{k}\boldsymbol{E}_0) &= 0, & \mu(\boldsymbol{k}\boldsymbol{H}_0) &= 0. \end{aligned} \tag{6.227}$$

Hereinafter we shall assume ε and μ not to turn into zero. Then from the last two equations it follows that the fields $\boldsymbol{E}_0$ and $\boldsymbol{H}_0$

are perpendicular to the wave vector $\boldsymbol{k}$. Such waves are referred to as transverse waves. We shall not discuss the longitudinal waves arising when, for example, $\varepsilon = 0$. (Then it is easy to see that $\boldsymbol{E}_0 \parallel \boldsymbol{k}$).

Multiplying vectorwise the first equation of (6.227) by the wave vector $\boldsymbol{k}$, we obtain

$$[\boldsymbol{k}[\boldsymbol{k}\boldsymbol{E}_0]] = \mu\omega[\boldsymbol{k}\boldsymbol{H}_0]. \tag{6.228}$$

Now the value of $[\boldsymbol{k}\boldsymbol{H}_0]$ can be taken from the second equation of (6.227), and the double vector product is written out with the help of the well-known formula

$$[\boldsymbol{k}[\boldsymbol{k}\boldsymbol{E}_0]] = \boldsymbol{k}(\boldsymbol{k}\boldsymbol{E}_0) - k^2\boldsymbol{E}_0 = -k^2\boldsymbol{E}_0.$$

In the last equation we took account of the transverse character of the field $\boldsymbol{E}_0$, i.e. the relation $(\boldsymbol{k}\boldsymbol{E}_0) = 0$. As a result of these transformations we obtain

$$(k^2 - \varepsilon\mu\omega^2 - i\sigma\omega\mu)\,\boldsymbol{E}_0 = 0. \tag{6.229}$$

Consequently, the amplitude $\boldsymbol{E}_0$ of a transverse electromagnetic wave in a stationary conducting medium can differ from zero only if the following equation is satisfied:

$$k^2 - \varepsilon\mu\omega^2 - i\sigma\mu\omega = 0. \tag{6.230}$$

This condition is referred to as a *dispersion relation*. One can readily see that the same dispersion relation (6.230) should be satisfied for the magnetic vector $\boldsymbol{H}_0$ in a transverse wave to differ from zero. In order to show this, one should multiply vectorwise the second relation of (6.227) by $\boldsymbol{k}$ and then make use of the first relation.

Let us assume that the wave frequency ω is a given quantity and the field depends only on the coordinate z and on time. Then we can represent the fields $\boldsymbol{E}$ and $\boldsymbol{H}$ in the form

$$\boldsymbol{E} = \boldsymbol{E}_0 e^{ikz - i\omega t}, \quad \boldsymbol{H} = \boldsymbol{H}_0 e^{ikz - i\omega t}. \tag{6.231}$$

It follows from (6.227) that we can assume the vector $\boldsymbol{E}_0$ to be directed in the positive direction of the x axis and the vector $\boldsymbol{H}_0$ in the positive direction of the y axis. Hence the three vectors $\boldsymbol{k}$, $\boldsymbol{E}_0$, $\boldsymbol{H}_0$ form the right-hand triad of vectors.

In the case of a fixed frequency ω the dispersion equation (6.230) yields the following values for the wave vector k:

$$k_{1,2} = \pm\,\omega\sqrt{\varepsilon\mu + i\,\frac{\sigma\mu}{\omega}}\,. \tag{6.232}$$

In a conducting medium the wave vector k turns out to be a complex quantity. Hereinafter it will be important that the conductivity σ is always positive. This can be deduced even from the

fact that Joule heat generated within a unit of volume of a conducting medium per unit of time is equal to $Q = j_{cond}E = \sigma E^2 > > 0$

Let us assume the frequency ω to be a positive quantity. Then in Eq. (6.232) the imaginary part of the radicand is also positive. Assuming that permeabilities ε and μ are also positive, we find that the solution k_1 is located in the first quadrant of the plane of imaginary variables, i.e. the imaginary and real parts of the solution k_1 are positive:

$$k_1 = k_1' + ik_1'' \quad (k_1', k_1'' > 0). \tag{6.233}$$

The second solution k_2 differs from the first one only by sign:

$$k_2 = -k_1 = -k_1' - ik_1'', \tag{6.234}$$

and we can present both solutions by means of the single equation

$$k = \pm k' \pm ik'', \tag{6.235}$$

where k' and k'' are positive quantities. The quantity $\pm k'$ is the real part of the wave vector k while the quantity $\pm k''$ is its imaginary part.

Substituting Eq. (6.235) in Eq. (6.231), we obtain

$$\boldsymbol{E} = \boldsymbol{E}_0{}^{\mp k''z} \cdot e^{i(\pm k'z - \omega t)}, \quad \boldsymbol{H} = \boldsymbol{H}_0 e^{\mp k''z} \cdot e^{i(\pm k'z - \omega t)}. \tag{6.236}$$

In these formulae one should take either all upper signs in the exponent, or all lower signs (wherever there is a choice between "$\pm$" and "$\mp$"). Thus we obtain the two solutions for the fields $\boldsymbol{E}$, $\boldsymbol{H}$: one proportional to

$$e^{-k''z} \cdot e^{i(k'z - \omega t)}, \tag{6.237}$$

and the other proportional to

$$e^{k''z} \cdot e^{-i(k'z + \omega t)}. \tag{6.238}$$

At first glance it seems that one of them, Eq. (6.237), attenuates exponentially as z increases (the factor $e^{-k''z}$), while the second one, Eq. (6.238), grows exponentially (the factor $e^{k''z}$). In actual fact, both solutions represent damping waves. To make sure that this is so, let us consider, for example, Eq. (6.238). It can be treated as a wave of the type

$$e^{-i(k'z + \omega t)}, \tag{6.239}$$

whose amplitude is equal to $e^{k''z}$, i.e. grows exponentially with z. It should be borne in mind that such a representation is valid only if the imaginary part k'' of the wave vector k is small compared to the real part k', i.e. the wave amplitude changes little at distances of the order of its wavelength.

The phase of the wave (6.239) is defined by the expression entering into the exponent:

$$\varphi = k'z + \omega t. \tag{6.240}$$

This phase is constant at

$$z = -\frac{\omega}{k'}t + \text{const}, \tag{6.241}$$

i.e. constant phase planes move along the axis at the velocity

$$v_{ph} = -\omega/k'. \tag{6.242}$$

This velocity v_{ph} is referred to as a *phase velocity* of a wave and determines the wave propagation direction. The minus sign in Eq. (6.242) means that the wave (6.238) propagates in the negative direction of the z axis. However, if we move in the negative direction of the z axis together with the wave, its amplitude, being proportional to the factor $e^{k''z}$, will attenuate exponentially. It can be seen that the wave (6.237) propagates in the positive direction of the z axis at the same (in magnitude) phase velocity. The amplitude of that wave is proportional to the factor $e^{-k''z}$ and, consequently, also attenuates in the propagation direction.

Thus, in a conducting medium there are two waves of a given frequency propagating in opposite directions and possessing phase velocities of equal magnitude. Amplitudes of each of these waves attenuate exponentially in the propagation direction.

It follows from the Maxwell equations that the amplitudes $\boldsymbol{E}_0$ and $\boldsymbol{H}_0$ are interrelated. This relationship can be expressed, for example, by the first equation (6.227). Taking this equation into account, we can write down the expressions for the fields (6.231) in the form

$$\begin{aligned} \boldsymbol{E} &= \boldsymbol{E}_0 e^{\mp k''z} \cdot e^{i(\pm k'z - \omega t)}, \\ \boldsymbol{H} &= \frac{1}{\mu\omega}[\boldsymbol{k}\boldsymbol{E}_0]\, e^{\mp k''z} \cdot e^{i(\pm k'z - \omega t)}. \end{aligned} \tag{6.243}$$

Thus, in order to determine a free plane electromagnetic wave in a conducting medium, one has to specify not only the characteristics of a medium ε, μ, σ, but also the field polarization, i.e. the direction and amplitude of the electric field $\boldsymbol{E}_0$.

From the solution of the dispersion equation (6.230) (see Eq. (6.232)) it is seen that the wave vector $\boldsymbol{k}$ is a complex quantity. Since we consider a unidimensional case when the field depends only on the coordinate z and time, we can assume the wave vector $\boldsymbol{k}$ to be directed along the z axis:

$$\boldsymbol{k} = \boldsymbol{k}' + i\boldsymbol{k}'', \tag{6.244}$$

where both vectors $\boldsymbol{k}'$ and $\boldsymbol{k}''$ are directed along the z axis. Then according to the third relation (6.227) the vector $\boldsymbol{E}_0$ is perpendicular to the z axis. Let this vector be directed along the x axis:

$$\boldsymbol{E}_0 = (E_0, 0, 0). \tag{6.245}$$

Then

$$\boldsymbol{H}_0 = (0, H_0, 0), \tag{6.246}$$

$$H_0 = \frac{1}{\mu\omega} kE_0, \tag{6.247}$$

i.e. the magnetic field is directed along the y axis as it is seen from Eq. (6.243).

For the sake of simplicity let us assume the quantity E_0 to be real. Then H_0 is a complex quantity since the expression for H_0 (6.247) includes the complex quantity k (6.232).

Let us represent the wave vector k in the form

$$k = k' + ik'' = |k| e^{i\varphi}, \tag{6.248}$$

where $|k| = \sqrt{(k')^2 + (k'')^2}$, $\tan\varphi = k''/k'$. Then the field equations (6.243) can be rewritten as

$$\begin{aligned} E_x &= E_0 e^{-k''z} e^{i(k'z-\omega t)}, \\ H_y &= \frac{1}{\mu\omega} |k| E_0 e^{-k''z} e^{i(k'z-\omega t+\varphi)}. \end{aligned} \tag{6.249}$$

Here we have used only the upper signs in Eq. (6.243), to make things simpler.

It is seen from these equations that in a conducting medium the waves of the electric and magnetic fields are displaced in phase by the angle $\varphi = \arctan (k''/k')$. When conductivity is absent, $k'' = 0$ and the phase displacement disappears.

The real physical fields $\boldsymbol{E}$ and $\boldsymbol{H}$ cannot be complex quantities so that a physical meaning may be ascribed either to real or to imaginary parts of Eq. (6.249). Taking, for example, the real parts of these expressions, we get

$$\begin{aligned} E_x &= E_0 e^{-k''z} \cos (k'z - \omega t), \\ H_y &= \frac{1}{\mu\omega} |k| E_0 e^{-k''z} \cos (k'z - \omega t + \varphi). \end{aligned} \tag{6.250}$$

The imaginary parts of Eq. (6.249) also yield the equivalent solutions:

$$\begin{aligned} E_x &= E_0 e^{-k''z} \sin (k'z - \omega t), \\ H_y &= \frac{1}{\mu\omega} |k| E_0 e^{-k''z} \sin (k'z - \omega t + \varphi). \end{aligned} \tag{6.251}$$

In conclusion, let us analyse Eq. (6.232) for the wave vector k in cases of low and high conductivity of a medium. Let us write

down Eq. (6.232), having chosen the plus sign in it for the sake of simplicity:

$$k=\omega\sqrt{\varepsilon\mu+i\frac{\sigma\mu}{\omega}}. \tag{6.252}$$

If the absolute value of the second term of the radicand is much less than of the first one (low conductivity), the following approximate expression is valid:

$$k=\omega\sqrt{\varepsilon\mu}+i\sqrt{\frac{\mu}{\varepsilon}}\frac{\sigma}{2}=k'+ik''. \tag{6.253}$$

In this case

$$k'=\omega\sqrt{\varepsilon\mu},\quad k''=\sqrt{\frac{\mu}{\varepsilon}}\frac{\sigma}{2}. \tag{6.254}$$

It is seen from these formulae that in the case of low conductivity the wave (6.250) attenuates e times over the distance L which is inversely proportional to conductivity:

$$L=\frac{1}{k''}=\sqrt{\frac{\varepsilon}{\mu}}\frac{2}{\sigma}. \tag{6.255}$$

If ε, μ and σ do not depend on frequency, the quantity L has the same magnitude for waves of all frequencies.

In the opposite extreme case of high conductivity we may ignore the first term of the radicand of Eq. (6.252) as compared to the second one. This yields

$$k=(1+i)\sqrt{\frac{\sigma\mu\omega}{2}}. \tag{6.256}$$

In this case the imaginary and real parts of the wave vector k are equal in magnitude. The distance L over which the wave attenuates e times is equal to

$$L=\frac{1}{k''}=\sqrt{\frac{2}{\sigma\omega\mu}}. \tag{6.257}$$

This quantity is referred to as a skin layer depth; the term arose due to the fact that a plane electromagnetic wave falling on a highly conducting body (metal) attenuates drastically so that the field differs from zero only in a thin surface layer, the depth of this layer being of the same order of magnitude as L. If σ and μ do not depend on frequency, the skin layer depth is inversely proportional to the square root of the incident wave frequency.

In the foregoing reasonings we assumed an electromagnetic wave frequency ω to be an assigned quantity and determined the wave vector from the dispersion equation (6.230). We could proceed otherwise by assigning a wavelength or its corresponding real wave vector $k=2\pi/\lambda$, and by calculating a wave frequency ω ex-

pressed via the wave vector k and the parameters of a medium ε, μ, σ:

$$\omega_{1,2} = -i\frac{\sigma}{\varepsilon} \pm \sqrt{\frac{4k^2}{\varepsilon\mu} - \frac{\sigma^2}{\varepsilon^2}}. \tag{6.258}$$

We shall not analyse this expression at length; note, however, that both solutions $\omega_{1,2}$ always have the negative imaginary part in the case of positive ε, μ, and σ, implying the wave attenuation with time. Indeed, the time dependence of the field has the form $e^{-i\omega t}$. The quantity ω is defined by the complex expression (6.258) which can be written as $\omega = \omega' + i\omega''$ where ω' is the real part of frequency and ω'' its imaginary part. When $\omega'' < 0$ one can easily see that the factor $e^{-i\omega t}$ attenuates with time as $e^{+\omega'' t}$. Do not forget that ω'' is a negative quantity while time varies only in a positive direction!

Let us consider now the equations describing potentials of an electromagnetic field in a moving conducting medium. It is known that in a stationary uniform isotropic conducting medium the conductivity current j_{cond} and the electric field E are interrelated:

$$j_{cond} = \sigma E.$$

When passing to four-dimensional designations, we must assume the current components to form a four-dimensional vector and the electric field to be expressed via the elements of a tensor of the second rank F_{ik} (see Eq. (6.29)). Therefore, in the relativistic invariant notation the quantity σ must be expressed via the elements of a certain tensor of the third rank:

$$j_{m,\,cond} = \sigma_{mkl} F_{kl}. \tag{6.259}$$

One can easily check that the tensor of the third rank σ_{mkl} can be expressed as follows:

$$\sigma_{mkl} = \frac{\sigma}{2}(\delta_{mk}U_l - \delta_{ml}U_k). \tag{6.260}$$

Substituting this expression in Eq. (6.259), we obtain $j_{cond} = \sigma E$ in the frame where a medium is stationary ($U_1 = U_2 = U_3 = 0$, $U_4 = ic$). If a medium moves at the velocity V, we get from Eqs. (6.259) and (6.260), and taking into account Eq. (6.29a):

$$j_{l,\,cond} = \sigma U_k F_{lk}. \tag{6.261}$$

In accordance with Eq. (6.81) this relation is equivalent to the following formulae:

$$j_{cond} = \sigma\Gamma\{E + [VB]\}, \quad \rho_{cond} = \sigma\Gamma\left(\frac{V}{c^2}E\right).$$

The first of these relations has a straightforward physical meaning: this is Ohm's law for a moving conductor. The factor at σ in the first relation defines the electric field in a stationary medium frame The second relation indicates that if a stationary conductor carrying a current is electrically neutral, an electric charge appears on this conductor when it moves at the velocity V (see § 6 1 for the physical interpretation of this phenomenon).

In the case of a moving conducting medium Eq. (6.60) should be written down as follows:

$$\frac{\partial f_{mk}}{\partial x_k} = (s_m + s_{m,\,cond}),$$

where $s_{m,\,cond} \equiv j_{m,\,cond}$ is defined by Eqs. (6.259), (6.260) and (6.261). Taking into account Eq. (6.261), the last equation is rewritten as

$$\frac{\partial f_{mk}}{\partial x_k} - \sigma F_{mk} U_k = s_m.$$

Let us substitute into this equation the quantity f_{mk} from Eq. (6.200) in which the tensor ε_{mkln} is defined from Eq. (6.201). Then the last equation becomes the equation describing the components of the tensor F_{mk}. If we now express F_{mk} in the obtained equation through the field potentials according to Eq. (6.199), we shall obtain the equation for field potentials in a moving conducting medium:

$$\left\{ \frac{\partial^2}{\partial x_k^2} - \frac{\varkappa}{c^2} \left(U_k \frac{\partial}{\partial x_k} \right)^2 - \sigma\mu \left(U_k \frac{\partial}{\partial x_k} \right) \right\} \Phi_m =$$
$$= -\mu c \left(\delta_{mk} + \frac{\varkappa c^{-2}}{1+\varkappa} U_m U_k \right) s_k. \quad (6.262)$$

Here the potentials Φ_m satisfy the following additional stipulation:

$$\frac{\partial \Phi_k}{\partial x_k} - \frac{\varkappa}{c^2} U_k U_l \frac{\partial \Phi_l}{\partial x_k} - \sigma\mu U_k \Phi_k = 0.$$

This stipulation is generalized from Eq. (6.210) to include the case of a conducting medium.

When the sources s_k are absent in a medium, we get the following system of uniform equations from Eq. (6.262):

$$\left\{ \frac{\partial^2}{\partial x_k^2} - \frac{\varkappa}{c^2} \left(U_k \frac{\partial}{\partial x_k} \right)^2 - \sigma\mu \left(U_k \frac{\partial}{\partial x_k} \right) \right\} \Phi_l = 0.$$

This system defines propagation of free electromagnetic waves in a moving medium which in a stationary frame has a dielectric permittivity ε, magnetic permeability μ and conductivity σ.

If a plane electromagnetic wave of the type (6.216) propagates in such a medium, the correlation between the frequency ω and wave vector $\boldsymbol{k}$ of that wave takes the form

$$\boldsymbol{k}^2 - \frac{\omega^2}{c^2} - \varkappa\Gamma^2(\omega - \boldsymbol{k}\mathbf{V})^2 + i\sigma\mu\Gamma(\omega - \boldsymbol{k}\boldsymbol{V}) = 0. \qquad (6.263)$$

This dispersion equation follows from the foregoing differential equation if one takes into account that the gradient operator $\nabla = \partial/\partial \boldsymbol{r}$ is equivalent to multiplication by $-i\boldsymbol{k}$, when applied to a plane wave of the type (6.216), and the operator of differentiating with respect to time $\partial/\partial t$ is equivalent to multiplication by $i\omega$. When the conductivity σ of a medium turns into zero, Eq. (6.263) passes into Eq. (6.218).

Having assigned the frequency ω and propagation direction of an electromagnetic wave, we can define the magnitude of the wave vector k (and thereby the wavelength $\lambda = 2\pi/k$) from Eq. (6.263). On the other hand, having assigned the magnitude and direction of the wave vector k, we can determine the wave frequency ω. If one of these quantities, k or ω, is assumed to be given, Eq. (6.263) becomes a quadratic equation with respect to the other quantity. This quadratic equation has complex coefficients, and therefore its solutions are also complex. It follows from the appearance of the wave (6.216) that if its frequency is complex, the wave is no more monochromatic and it either grows or attenuates exponentially with time. In this case an attenuation (or growth) index is equal to the imaginary part of the frequency ω. When the imaginary part ω'' of the frequency ω is positive, the wave attenuates with time, and when it is negative, the wave grows with time.

When in Eq. (6.263) the frequency ω and propagation direction of a wave, that is the angle θ between $\boldsymbol{V}$ and $\boldsymbol{k}$, are assigned, we obtain a quadratic equation for the absolute value of the wave vector k. The solution of this equation yields, generally speaking, complex values of k. In this case two conjugate complex roots of Eq. (6.263) correspond to the exponential growing or attenuation of the wave in space. From here on we shall confine ourselves to the case of low attenuation when the imaginary part of the solution of Eq. (6.263) for k can be considered small as compared to the real part. In this particular case the sign of the imaginary part of k does not define whether the wave grows or attenuates in space. Indeed, let one of the solutions of Eq. (6.263), with ω and θ assigned, be equal to $k' + ik''$, where k' is a real and k'' an imaginary part of k. The wave vector direction is defined by the unit vector $\boldsymbol{n}$ so that

$$\boldsymbol{k} = k\boldsymbol{n} = (k' + ik'')\,\boldsymbol{n}.$$

Let us direct the z axis of the Cartesian system of coordinates along the vector $\boldsymbol{n}$. Then the wave (6.216) is written down as

$$A = A_0 e^{i(\omega t - \boldsymbol{kr})} = A_0 e^{k''z} \cdot e^{i(\omega t - k'z)}. \tag{6.264}$$

When the attenuation is low ($k'' \ll k'$), Eq. (6.264) can be considered to define a wave possessing the wave vector k' and frequency ω, with its amplitude varying according to the exponential law $e^{k''z}$. Suppose k'' is a positive quantity. To make any conclusions concerning the wave behaviour, one needs to know the wave propagation direction, i.e. the sign of its phase velocity. The phase velocity of the wave is equal to the ratio ω/k'. Indeed, the plane of a constant phase of the wave (6.264) is defined by the relation $\omega t - k'z = \text{const}$, whence $z = \frac{\omega}{k'} t - \frac{\text{const}}{k'}$. It is seen from the last relation that the plane of a constant phase travels at the velocity ω/k'. When $\omega/k' > 0$, the wave (6.264) propagates in the positive direction of the z axis. Then if $k'' > 0$, the wave grows, and if $k'' < 0$, it attenuates. And when $\omega/k' < 0$, the wave propagates in the negative direction of the z axis. Then if $k'' > 0$, the wave attenuates in its propagation direction (although its amplitude does grow in the positive direction of the z axis). Thus, to determine whether the wave grows or attenuates, it is not sufficient to know the law according to which the wave amplitude varies in space; one has to know the wave propagation direction as well.

There is a simple method making it possible to find out whether the wave grows or attenuates in the direction of its propagation. Let us analyse the expression $\omega k''/k'$. When it is positive, the wave grows in the direction of its propagation; in the opposite case the wave attenuates. It can be easily seen that the expression $\omega k''/k'$ is the product of the phase velocity of the wave by the decrement of its attenuation in space.

Let us consider now the solution of the dispersion equation (6.263). Let the magnitude of the wave vector be equal to k and its direction form the angle θ with the velocity vector $\boldsymbol{V}$ of a medium. In this case $\boldsymbol{kV} = kV\cos\theta$. The dispersion equation (6.263) is a quadratic equation with respect to the frequency ω. Solving it for the case of low conductivity σ and discarding all degrees of σ exceeding the first, we get

$$\omega_{1,2} = (1+\varkappa\Gamma^2)^{-1}\left\{\left[\varkappa\Gamma^2 kV\cos\theta \pm ck\sqrt{\Delta} + i\frac{c}{2}\sigma\mu\Gamma\left[1 \mp \frac{\mathrm{B}\cos\theta}{\sqrt{\Delta}}\right]\right\}, \tag{6.265}$$

where $\varkappa = \frac{n^2-1}{c^2}$, $\mathrm{B} = \frac{V}{c}$, $\Gamma = (1-\mathrm{B}^2)^{-1/2}$, and $\Delta = 1 + \varkappa\Gamma^2(1 - \mathrm{B}^2\cos^2\theta)$. This equation shows that if in the frame of a sta-

tionary medium $\varepsilon\mu > \varepsilon_0\mu_0$, i.e. $\varkappa = (n^2 - 1)/c^2 > 0$, the imaginary part of the frequency ω is always positive for both solutions, whatever the velocity $\boldsymbol{V}$ of motion of a medium. This means that for a given wave vector $\boldsymbol{k}$ the wave (6.216) always attenuates with time. The attenuation decrement is proportional to the conductivity σ.

Now let us consider the case when the given characteristics of the wave (6.216) are the frequency ω and the wave propagation direction defined by the angle θ. Then from the dispersion equation (6.263) one can determine the magnitude of the wave vector k corresponding to the given values of ω and θ. The solutions of this equation have the form $k_{1,2} = k'_{1,2} + ik''_{1,2}$. In the case of small σ we obtain after minor transformations

$$\begin{aligned} ck'_1 &= \omega(1 + \varkappa\Gamma^2)\left(\varkappa B\Gamma^2 \cos\theta + \sqrt{\Delta}\right)^{-1}, \\ 2ck''_1 &= -c\sigma\mu\Gamma(1 - B^2\cos^2\theta)\left(1 + B\cos\theta\sqrt{\Delta}\right)^{-1}; \\ ck'_2 &= -\omega\left(\varkappa B\Gamma^2\cos\theta + \sqrt{\Delta}\right)(1 - \varkappa B^2\Gamma^2\cos^2\theta)^{-1}, \\ 2ck''_2 &= c\sigma\mu\Gamma\left(1 + B\cos\theta\sqrt{\Delta}\right)(1 - \varkappa B^2\Gamma^2\cos^2\theta)^{-1}\Delta^{-1/2}. \end{aligned} \tag{6.266}$$

Here the quantity Δ is always positive due to the assumptions made earlier. Using Eq. (6.266) one can obtain expressions for $\omega k''_{1,2}/k'_{1,2}$. They have the following form:

$$\frac{\omega}{k'_1} k''_1 = -\frac{\sigma\mu c^2\Gamma(1 - B^2\cos^2\theta)\left(\varkappa B\Gamma^2\cos\theta + \sqrt{\Delta}\right)}{2(1 + \varkappa\Gamma^2)\left(1 + B\cos\theta\sqrt{\Delta}\right)},$$

$$\frac{\omega}{k'_2} k''_2 = -\frac{\sigma\mu c^2\Gamma\left(1 + B\cos\theta\sqrt{\Delta}\right)}{2\left(\varkappa B\Gamma^2\cos\theta + \sqrt{\Delta}\right)\sqrt{\Delta}}.$$

It is immediately seen from this that if $\varkappa > 0$, the products $\omega k''_{1,2}/k'_{1,2}$ are always negative, whatever the velocity of motion of a medium. This means that in a moving conducting medium the wave (6.216) always attenuates in the direction of its propagation. The only characteristic property of a moving medium is that when the velocity of motion of a medium satisfies the condition $1 - \varkappa B^2\Gamma^2\cos^2\theta = 0$ or $B = 1/(1 + \varkappa\cos^2\theta)$, the real and imaginary parts of the second solution change sign simultaneously.

The potentials obtained by solving Eqs. (6.211) and (6.262) can be successfully used to solve other problems (see the bibliography at the end of the book), most of which, however, lie outside the scope of this book.

CHAPTER 7

OPTICAL PHENOMENA AND THE SPECIAL THEORY OF RELATIVITY

Light, being a special case of electromagnetic waves, is described by the Maxwell theory. As we have seen in the foregoing chapter, the Maxwell theory meets all requirements of the theory of relativity, and therefore must accurately describe the properties of such a typical relativistic object as light. But even in the theory of relativity the propagation of light *in vacuo* holds a special position. We have already pointed out that the velocity of light *in vacuo* is the ultimate feasible velocity of signal transmission and an unattainable velocity limit for objects possessing a finite rest mass. Besides, at the basis of the STR lies the statement about the velocity of light *in vacuo* being the same in all IFRs.

The Maxwell theory is a macroscopic theory. In this chapter it is convenient to examine a microscopic approach and to some extent even quantum mechanical methods. We mean the introduction of photons here. In some respect, the introduction of quantum concepts leads to a very descriptive picture. The utilization of the theory of relativity becomes indispensable when we consider optical phenomena associated with a relative motion of bodies (the Doppler effect, aberration).

§ 7.1. Properties of plane light waves. The Maxwell theory shows that in a uniform isotropic medium ($\varepsilon = \text{const}$, $\mu = \text{const}$) whose conductivity σ is equal to zero, the time-dependent field vectors $\boldsymbol{E}$ and $\boldsymbol{H}$ (as well as $\boldsymbol{D}$ and $\boldsymbol{B}$ which are proportional to them) satisfy the wave equations

$$\begin{aligned} \Box\, \boldsymbol{E} &\equiv \Delta \boldsymbol{E} - \frac{1}{v^2}\frac{\partial^2 \boldsymbol{E}}{\partial t^2} = 0, \\ \Box\, \boldsymbol{H} &\equiv \Delta \boldsymbol{H} - \frac{1}{v^2}\frac{\partial^2 \boldsymbol{H}}{\partial t^2} = 0. \end{aligned} \tag{7.1}$$

This signifies that in a uniform non-conducting medium the waves can propagate, whose phase velocity $v = 1/\sqrt{\varepsilon\mu}$ is defined exclusively by the properties of a medium. One of the possible solutions of Eqs. (7.1) yields the plane waves:

$$\boldsymbol{E} = \boldsymbol{E}_0 e^{i(\omega t - \boldsymbol{k}\boldsymbol{r})}, \quad \boldsymbol{H} = \boldsymbol{H}_0 e^{i(\omega t - \boldsymbol{k}\boldsymbol{r})}, \tag{7.2}$$

where ω is the circular frequency. Here the field vectors are

assumed to depend harmonically on time, and the wave vector to be directed along a normal to the surface of equal phases (the wave front). It follows from Eqs. (7.1) that the absolute value of the wave vector $\boldsymbol{k}$ is equal to ω/v provided waves propagate in a medium and to ω/c when they propagate *in vacuo*. Since $n = \sqrt{\varepsilon\mu/\varepsilon_0\mu_0}$, one can write in a general case

$$\boldsymbol{k} = \frac{\omega}{c} n\boldsymbol{s}, \tag{7.3}$$

where $\boldsymbol{s}$ is a unit vector oriented along the propagation direction. In the case of vacuum $n = 1$. The phase of the wave (7.2) is equal to $\omega t - \boldsymbol{kr}$ and therefore the surface of equal phases is defined by the equation $\omega t - \boldsymbol{kr} = \text{const}$. At a given moment of time it represents the plane $\boldsymbol{kr} = \text{const}$, with the vector of its normal directed along $\boldsymbol{k}$ ($\boldsymbol{r}$ is a conventional three-dimensional radius vector). In the course of time this plane translates in space parallel to itself in accordance with the equation $\boldsymbol{kr} = \text{const} + \omega t$.

The plane waves (7.2) must satisfy not only Eqs. (7.1) but also the Maxwell equations (6.56) and (6.57) in the absence of charges ($\rho = 0$) and currents ($\boldsymbol{j} = 0$); substituting Eq. (7.2) into the Maxwell equations we obtain the following results. In a plane wave propagating in a uniform medium the vectors $\boldsymbol{E}$, $\boldsymbol{H}$ and $\boldsymbol{k}$ form a clockwise triad, i.e. they are mutually perpendicular and the vector cross product of any pair of them, taken in the order indicated, defines the direction of the third vector.

As to the relationship between the amplitudes, the following equation is valid: $\sqrt{\mu}H = \sqrt{\varepsilon}E$. Consequently, in the case of vacuum, when $\boldsymbol{B} = \mu_0\boldsymbol{H}$ and $\varepsilon = \varepsilon_0$, we get $E = cB$.

The direction of the Poynting vector $\boldsymbol{S}$ coincides with that of the vector $\boldsymbol{k}$ while its absolute value is equal to the product of the energy density in the plane wave and the wave propagation velocity v, i.e. $S = wv$, or $\boldsymbol{S} = wv(\boldsymbol{k}/k)$ where w is the energy density in the electromagnetic wave. This result has a clear physical meaning: the Poynting vector determines an energy flux across a unit of area oriented normally to the incident wave per unit of time. But the energy flowing across a unit of area per unit of time is contained within the cylinder whose directrix is formed by the contour of that unit of area and generatrix by straight lines parallel to the wave propagation direction. The cylinder's height should be taken equal to v. In this case the quantity v defines the volume of the cylinder thus formed, and the product vw, the electromagnetic field energy contained in the cylinder. All this results in $S = vw$. Note also that in a plane wave

$$w = \frac{\varepsilon E^2 + \mu H^2}{2} = \varepsilon E^2,$$

while *in vacuo* $w = \varepsilon_0 E^2$.

The momentum of a unit of volume (the momentum density) of an electromagnetic field *in vacuo* $\boldsymbol{g}$ is equal to $\boldsymbol{S}/c^2$. In the case of a plane wave *in vacuo* when $S = cw$, we get $\boldsymbol{g} = (w/c)\,(\boldsymbol{k}/k)$, whence

$$g = w/c. \tag{7.4}$$

The field momentum density in a medium will be examined in § 7 of this chapter. Recalling the invariants I_1 and I_2 of an electromagnetic field (§ 6.5), we find that in the case of a plane wave *in vacuo* both invariants turn to zero. This means that in any frame the vectors $\boldsymbol{E}$ and $\boldsymbol{H}$ of a plane wave are orthogonal, and the ratio of their amplitudes is always the same. In the frame K' a plane wave must take the form

$$\boldsymbol{E}' = \boldsymbol{E}'_0 e^{i(\omega' t' - \boldsymbol{k}'\boldsymbol{r}')}. \tag{7.5}$$

The phase of the wave at the world point $\vec{R}(\boldsymbol{r}, ict)$ cannot depend on the choice of a reference frame. Therefore, the phase $\omega t - \boldsymbol{kr}$ must be an invariant of the Lorentz transformation. Consequently,

$$\omega t - k_x x - k_y y - k_z z = \omega' t' - k'_x x' - k'_y y' - k'_z z'. \tag{7.6}$$

Substituting the transformation formulae for x', y', z' and t' from Eq. (2.37) into the right-hand side of Eq. (7.6), we obtain

$$\omega t - k_x x - k_y y - k_z z = \omega' \Gamma \left(t - \frac{B}{c} x\right) - k'_x \Gamma (x - Vt) - k'_y y - k'_z z.$$

This is an identity with respect to t, x, y, z. Taking into account that $k = \omega/c$ and $k_x = \frac{\omega}{c} s_x$, $k_y = \frac{\omega}{c} s_y$, $k_z = \frac{\omega}{c} s_z$ ($\boldsymbol{s}$ is a unit vector whose direction coincides with that of $\boldsymbol{k}$), we get

$$\omega = \omega' \Gamma (1 + Bs'_x), \quad \omega s_x = \omega' \Gamma (B + s'_x), \quad \omega s_y = \omega' s'_y, \quad \omega s_z = \omega' s'_z. \tag{7.7}$$

Here $\boldsymbol{k}' = \frac{\omega'}{c} \boldsymbol{s}'$.

From these equations one can easily obtain the formulae describing the Doppler effect, that is the light wavelength variation when emitted by a source moving relative to an observer, and an aberration of light, that is the change in the direction of a light beam on transition from one inertial frame of reference to another. To eliminate reiteration, however, we shall derive Eqs. (7.7) in a somewhat different fashion and then, in the next section, investigate their consequences.

We shall take the four-dimensional approach from the very beginning. It has been already pointed out that the phase $\omega t - \boldsymbol{kr}$

must be an invariant of the Lorentz transformation. But this expression becomes an invariant automatically if it is represented as a scalar product of 4-vectors (the scalar product invariance is shown in Appendix I, § 4). For this purpose it is sufficient to introduce the 4-wave vector $\vec{k}\left(\boldsymbol{k}, i\frac{\omega}{c}\right)$ along with the 4-radius vector $\vec{R}$ $(\boldsymbol{r}, ict)$. Then, $\omega t - \boldsymbol{kr} = -\vec{k}\vec{R}$. The introduction of the 4-wave vector $\vec{k}$ is convenient mainly because we immediately obtain the rule for its component transformation on transition from one IFR to another. A plane light wave propagating in the frame K' changes its direction and observed frequency on transition to the frame K. We shall see that an amplitude of a plane wave also varies. The transformation equations for quantities characterizing a light wave in the reference frames K and K' can be readily obtained if one takes into consideration that in a plane light wave the conventional wave vector $\boldsymbol{k}$, together with $i(\omega/c)$, forms the 4-vector $\vec{k}$.

§ 7.2. A 4-wave vector. The Doppler effect. Aberration of light. Let us consider a plane light wave observed in the reference frame K' and described by the 4-vector $\vec{k}'$. The frame K' is chosen so that the light beam propagates in it in the plane (x', y') at the angle θ' to the x' axis. Write out the 4-vector components:

$$\begin{aligned} &k_1' = k'\cos\theta' = \frac{\omega'}{c}\cos\theta', \quad && k_2' = k'\sin\theta' = \frac{\omega'}{c}\sin\theta', \\ &k_3' = 0, && k_4' = i\frac{\omega'}{c} = ik'. \end{aligned} \tag{7.8}$$

Now let us find the components of the 4-vector $\vec{k}$ in the frame K. In accordance with the general equations (4.10a)

$$k_1 = \Gamma(k_1' - iBk_4'), \quad k_2 = k_2', \quad k_3 = k_3', \quad k_4 = \Gamma(k_4' + iBk_1'). \tag{7.9}$$

Since $k_3 = 0$, in the frame K the beam propagates in the plane (x, y) as well. Consequently, the 4-vector $\vec{k}$ has the components $\vec{k}\left(\frac{\omega}{c}\cos\theta, \frac{\omega}{c}\sin\theta, 0, i\frac{\omega}{c}\right)$ in the frame K. From the last formula (7.9) we find that

$$i\frac{\omega}{c} = \Gamma\left(i\frac{\omega'}{c} + iB\frac{\omega'}{c}\cos\theta'\right),$$

or

$$\omega = \omega'\frac{1 + B\cos\theta'}{\sqrt{1 - B^2}} = \omega'\Gamma(1 + B\cos\theta'). \tag{7.10}$$

Consequently, if in the frame K' the light frequency is equal to ω', it will be different in the frame K in accordance with Eq. (7.10) (cf. Eqs. (7.7)). It follows from the first formula (7.9) that

$$\frac{\omega}{c}\cos\theta = \Gamma\left(\frac{\omega'}{c}\cos\theta' - iBi\frac{\omega'}{c}\right),$$

or, if Eq. (7.10) is taken into account,

$$\cos\theta = \frac{\omega'}{\omega}\Gamma(\cos\theta' + B) = \frac{\cos\theta' + B}{1 + B\cos\theta'}. \tag{7.11}$$

The second formula (7.9), together with Eq. (7.10), yields

$$\sin\theta = \frac{\omega'}{\omega}\sin\theta' = \frac{\sqrt{1-B^2}}{1 + B\cos\theta'}\sin\theta' = \frac{\sin\theta'}{\Gamma(1 + B\cos\theta')}. \tag{7.12}$$

Using Eqs. (7.11) and (7.12), one can easily find the expression for $\sin\theta'$ in terms of the angle θ:

$$\sin\theta' = \frac{\sin\theta}{\Gamma(1 - B\cos\theta)}. \tag{7.12'}$$

Note that Eq. (7.12′) is immediately obtained from Eq. (7.12) by substituting unprimed quantities for primed ones and vice versa and by taking the velocity V with the opposite sign. The equations obtained make it possible to interpret quantitatively the two optical effects: the Doppler effect and aberration of light. The Doppler effect which is observed for waves of any kind consists in the fact that in the case of a relative motion of a source and an observer (receiver) the frequency (of sound or light) determined by the observer differs from that measured in the reference frame in which the source is at rest.

Let a source be at rest in the frame K'. Then the instruments resting in that frame will determine the natural frequency ω_0 of the light source ($\omega_0 = \omega'$).

When determining the frequency ω in the frame K, we need to convert the angle θ' into the angle θ. It follows from Eq. (7.11) that

$$\cos\theta' = \frac{\cos\theta - B}{1 - B\cos\theta},$$

whence $1 + B\cos\theta' = (1 - B^2)/(1 - B\cos\theta)$, and, consequently, Eq. (7.10) can be rewritten in the final form:

$$\omega = \omega_0\frac{\sqrt{1-B^2}}{1 - B\cos\theta}. \tag{7.13}$$

This is the equation describing the Doppler effect. An observer in the frame K will observe the radiation frequency ω differing from the natural frequency ω_0 of the source. The observed frequency ω depends not only on the relative velocity of the source and the observer ($B = V/c$), but also on the angle θ at which light comes to the observer.

In particular, if the radiation comes along the relative velocity direction, we observe the so-called *radial Doppler effect*. If the frame K' is to the right of K, the source moves away from the observer and light propagates in the direction opposite to the x axis

(see Fig. 7.1*a*). Consequently, $\cos\theta = \cos\pi = -1$. Then from Eq. (7.13) we obtain the frequency ω and period $T = 2\pi/\omega$:

$$\omega = \omega_0 \sqrt{\frac{1-\mathrm{B}}{1+\mathrm{B}}}, \quad T = T_0 \sqrt{\frac{1+\mathrm{B}}{1-\mathrm{B}}}.$$

An observer receiving light from a source moving away from him finds a frequency to decrease.

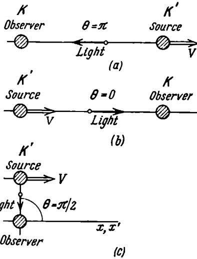


Fig. 7.1. The radial Doppler effect: (*a*) an observer and a source move away from each other; (*b*) an observer and a source draw together, (*c*) the transverse Doppler effect.

On the other hand, when the frame K' is to the left of K (see Fig. 7.1*b*), $\cos\theta = 1$ and a source approaches an observer:

$$\omega = \omega_0 \sqrt{\frac{1+\mathrm{B}}{1-\mathrm{B}}}, \quad T = T_0 \sqrt{\frac{1-\mathrm{B}}{1+\mathrm{B}}}.$$

The frequency of light received by an observer increases as compared to the natural frequency ω_0. To an accuracy of B^2 terms the last two formulae can be rewritten as follows (the easiest way is to multiply both the radicand numerator and denominator by the numerator):

$$\omega = \omega_0(1-\mathrm{B}), \quad \omega = \omega_0(1+\mathrm{B}).$$

Both formulae can be combined:

$$\frac{\omega - \omega_0}{\omega_0} \equiv \frac{\Delta\omega}{\omega_0} = \pm\,\mathrm{B}.$$

Thus, the radial Doppler effect proves to be an effect of the first order with respect to B. To an accuracy of the second order with respect to B the formulae obtained coincide with the classical ones following from the fundamental considerations (§ 3.3).

When light is observed in the direction which is perpendicular to the source velocity direction, i.e. $\theta = \pi/2$ (see Fig. 7.1*c*), we witness the so-called *transverse Doppler effect*. It is described by the formula

$$\omega = \omega_0 \sqrt{1 - B^2}$$

and depends on B^2 this time. If a source motion velocity is non-relativistic, the binomial expansion yields

$$\omega = \omega_0 (1 - B^2/2).$$

Since this is a second-order effect, its observation is much more difficult to perform as compared to the radial effect. No wonder, the transverse Doppler effect was observed as late as 1938 (Ives) when the relativistic formula was wholly confirmed *. We would like to point out here once more that according to the classical theory no radial Doppler effect should exist (cf. § 3.3). The radial Doppler effect arises only due to the relativity of time intervals between events.

Let us rewrite Eq. (7.13) in the form that was used in § 6.15. We shall group the quantities pertaining to the frame K on the right-hand side:

$$\omega_0 = \Gamma\omega\left(1 - \frac{V}{c}\cos\theta\right) = \Gamma\left(\omega - \frac{\omega}{c}V\cos\theta\right) = \Gamma(\omega - \boldsymbol{kV}). \quad (7.14)$$

The natural frequency appears on the left-hand side while the right-hand side contains the frequency observed in the reference frame moving at the velocity $\boldsymbol{V}$, the light propagation direction being defined by the vector $\boldsymbol{k}$.

Eqs. (7.11) and (7.12) coincide with the formulae derived directly from the velocity transformation formulae; therefore they fully describe the phenomenon of aberration that was mentioned in § 3.6.

In particular, the expression for an aberration angle follows from Eqs. (7.11) and (7.12):

$$\tan\theta = \frac{\sin\theta' \sqrt{1 - B^2}}{B + \cos\theta'}. \quad (7.15)$$

To conclude this section, let us derive the transformation formula for a solid angle element written in spherical coordinates.

* The details of Ives's experiments can be found in the monograph by Landsberg G. S., *Optics*, 1976, "Nauka" Publishing House (in Russian).

We shall orient the polar axis in the direction of relative motion of two frames (the x, x' axis). In the frame K' the solid angle element $d\Omega'$ is written down as $d\Omega' = \sin\theta'\, d\theta'\, d\varphi' = -d(\cos\theta')d\varphi'$. Since the y and z coordinates do not change, the φ coordinate, that is the projection on a plane perpendicular to the motion direction, does not change either: $\varphi = \varphi'$ and $d\varphi = d\varphi'$. From the formula preceding Eq. (7.13) it follows that

$$d(\cos\theta') = -\frac{1-\mathrm{B}^2}{(1-\mathrm{B}\cos\theta)^2}\sin\theta\, d\theta,$$

whence the sought for transformation formula is obtained:

$$d\Omega' = \frac{1}{\Gamma^2(1-\mathrm{B}\cos\theta)^2}\sin\theta\, d\theta\, d\varphi = \frac{d\Omega}{\Gamma^2(1-\mathrm{B}\cos\theta)^2}, \tag{7.16}$$

since in the frame K the solid angle element $d\Omega$ is equal to

$$d\Omega = \sin\theta\, d\theta\, d\varphi.$$

§ 7.3. A plane wave limited in space. The transformation of the plane wave energy and amplitude. Let us calculate the components of an energy-momentum-tension tensor for the case of a plane wave. We shall orient the x' axis along the wave propagation direction, the y' axis along the vector $\boldsymbol{E}'$ and the z' axis along the vector $\boldsymbol{B}'$. With such a choice of axes, $E'_x = D'_x = B'_x = H'_x = E'_z = D'_z = H'_y = B'_y = 0$. The tensor T'_{ik} takes the simple form (see Eqs. (6.128), (6.148) and (6.151)):

$$T'_{ik} = \begin{pmatrix} -w' & 0 & 0 & -iw' \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -iw' & 0 & 0 & w' \end{pmatrix}. \tag{7.17}$$

We shall also need the components of the tensor T'_{ik} in the case when a plane wave propagates in the (x', y') plane at the angle θ' to the x' axis. Such a transition is accomplished through a simple rotation of a coordinate system; the matrix of this coordinate transformation takes the form

$$\bar{a}_{ik} = \begin{pmatrix} \cos\theta' & -\sin\theta' & 0 & 0 \\ \sin\theta' & \cos\theta' & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{7.18}$$

Transforming the components of the tensor (7.17) through the use of the matrix (7.18) according to the general rules of tensor trans-

formation, we reach the tensor $\tilde{T}'_{ik}$:

$$\tilde{T}'_{ik}=\begin{pmatrix} -w'\cos^2\theta' & -w'\sin\theta'\cos\theta' & 0 & -iw'\cos\theta' \\ -w'\sin\theta'\cos\theta' & -w'\sin^2\theta' & 0 & -iw'\sin\theta' \\ 0 & 0 & 0 & 0 \\ -iw'\cos\theta' & -iw'\sin\theta' & 0 & w' \end{pmatrix}. \tag{7.19}$$

Therefore, specifically,

$$\tilde{w}'=w', \quad \tilde{S}'_x=cw'\cos\theta', \quad \tilde{T}'_{11}=-w'\cos^2\theta'. \tag{7.20}$$

Let us prove the following theorem: a plane wave limited in space along its propagation direction (such a wave is sometimes called "a train of waves") possesses a momentum and an energy making up a 4-vector similar to a 4-vector of energy-momentum of a material particle. (This theorem is a particular case of the more general theorem *.) To prove the theorem we have to know the formula defining a change in volume occupied by a train of waves on transition from one inertial frame to another. The difficulty arising here is caused by the fact that the train of waves moves at the velocity of light c so that the volume of the train cannot be measured in the proper frame of reference. It is impossible to introduce the reference frame moving at the velocity of light! However, one can bypass the introduction of the proper volume, having finally accomplished the limit transition to the velocity of light.

Let a certain volume move as a whole in the frame K' at the velocity v', with its value being equal to $\mathscr{V}_0$ in the proper frame of reference. Then according to Eq. (3.28)

$$\mathscr{V}'=\mathscr{V}_0\sqrt{1-\frac{v'^2}{c^2}}. \tag{7.21}$$

If one considers this volume in the frame K, its velocity v will be determined by Eq. (3.41), and, consequently, the magnitude $\mathscr{V}$ of this volume in the frame K will be equal to

$$\mathscr{V}=\mathscr{V}_0\sqrt{1-\frac{v^2}{c^2}}=\mathscr{V}_0\frac{\sqrt{1-\frac{v'^2}{c^2}}\sqrt{1-\frac{V^2}{c^2}}}{1+\frac{v'}{c}\mathrm{B}\cos\theta'}=\mathscr{V}'\frac{\sqrt{1-\mathrm{B}^2}}{1+\frac{v'}{c}\mathrm{B}\cos\theta'};$$

* The general theorem is presented in [13] § 57 and W Heitler's *The Quantum Theory of Radiation*, Oxford, 1954, § 2 This general theorem can be formulated as follows in the space region where the tensor T_{ik} satisfies the condition $\frac{\partial}{\partial x_k}T_{ik}=0$ and $T_{ik}=0$ at the boundaries, the components T_{4k} constitute the components of a 4-vector.

the second equality follows from Eq. (3.41). Here v' is the motion velocity in the frame K'. Now, passing to the limit $v' \to c$, we get the required formula:

$$\mathscr{V} = \mathscr{V}' \frac{\sqrt{1 - B^2}}{1 + B\cos\theta'}. \tag{7.22}$$

Thus, if a certain volume in the frame K' is equal to $\mathscr{V}'$ we shall observe in the frame K, moving at the velocity V relative to K', a volume whose magnitude is determined from Eq. (7.22). It is understood that there is a similar relation for volume differentials:

$$d\mathscr{V} = d\mathscr{V}' \frac{\sqrt{1 - B^2}}{1 + B\cos\theta'} = \frac{1}{\Gamma(1 + B\cos\theta')} d\mathscr{V}'. \tag{7.23}$$

Let us go back to proving the theorem. Applying the general tensor transformation formulae (A.I. 31) to the tensor (7.17), we shall obtain the fourth line components of the matrix T_{ik} in the following form:

$$\begin{aligned} &T_{41} = -i\Gamma^2 w' (1 + B)^2, \quad T_{42} = 0, \\ &T_{43} = 0, \quad T_{44} = \Gamma^2 w' (1 + B)^2, \end{aligned} \tag{7.24}$$

and for the tensor (7.19) the components

$$\begin{aligned} &T_{41} = -i\Gamma^2 w' (B + \cos\theta')(1 + B\cos\theta'), \\ &T_{42} = -i\Gamma w' \sin\theta' (1 + B\cos\theta'), \\ &T_{43} = 0, \quad T_{44} = \Gamma^2 w' (1 + B\cos\theta')^2. \end{aligned} \tag{7.25}$$

Naturally, when $\theta' = 0$, Eqs. (7.25) pass into Eqs. (7.24). Let us show now that the components (7.25) and, consequently, in the specific case, (7.24) as well, transform vectorwise in the necessary frame of reference when multiplied by the volume or the volume element. Indeed, for example,

$$\begin{aligned} &T_{41} d\mathscr{V} = \Gamma\left(-iw'\cos\theta' - iBw'\right) d\mathscr{V}' = \Gamma\left(T'_{41}\, d\mathscr{V}' - iBT'_{44}\, d\mathscr{V}'\right), \\ &T_{42}\, d\mathscr{V} = T'_{42}\, d\mathscr{V}', \quad T_{43}\, d\mathscr{V} = T'_{43}\, d\mathscr{V}' = 0, \\ &T_{41}\, d\mathscr{V} = \Gamma(w' + Bw'\cos\theta)\, d\mathscr{V}' = \Gamma(T'_{44}\, d\mathscr{V}' + iBT'_{14}\, d\mathscr{V}'). \end{aligned} \tag{7.26}$$

Comparing the obtained formulae (7.26) with the vector transformation formulae, we conclude that the quantities T_{41}, T_{42}, T_{43}, T_{44}, i.e. the fourth line components of the tensor (7.17) or (7.19) multiplied by the corresponding volume element make up a 4-vector.

Of course, this result holds after integrating over volume or multiplying by a total volume, provided the tensor components T_{ik} do not depend on coordinates as it is the case in a plane wave.

Let us find the total energy of a train of waves in the frame K' (see Eq. (7.19)):

$$U' = \int \tilde{T}'_{44}\, d\mathscr{V}' = \int w'\, d\mathscr{V}'. \tag{7.27}$$

The total momentum components of a train of waves are defined by the following formulae:

$$G'_x = \frac{i}{c}\int \tilde{T}'_{14}\, d\mathscr{V}' = \frac{i}{c}\int (-iw'\cos\theta')\, d\mathscr{V}' = \frac{U'}{c}\cos\theta',$$
$$G'_y = \frac{i}{c}\int \tilde{T}'_{24}\, d\mathscr{V}' = \frac{U'}{c}\sin\theta', \quad G'_z = 0. \tag{7.28}$$

Similar calculations can also be accomplished in the frame K. Eqs. (7.25) allow the total energy transformation to be performed directly:

$$U = \int T_{44}\, d\mathscr{V} = \Gamma^2 (1 + \mathrm{B}\cos\theta')^2 \int \frac{w'\, d\mathscr{V}'}{\Gamma(1 + \mathrm{B}\cos\theta')} =$$
$$= \Gamma(1 + \mathrm{B}\cos\theta')\, U'. \tag{7.29}$$

Calculating also the total momentum components

$$G_x = \frac{i}{c}\int T_{14}\, d\mathscr{V} = \frac{i}{c}(-i\Gamma)(\mathrm{B} + \cos\theta')\int w'\, d\mathscr{V}' =$$
$$= \frac{1}{c}\Gamma(\mathrm{B} + \cos\theta')\, U' = \frac{U}{c}\cos\theta, \tag{7.30}$$

we used Eq. (7.11) in the last equality; in much the same way, using Eq. (7.12), we get

$$G_y = \frac{i}{c}\int T_{24}\, d\mathscr{V} = \frac{1}{c}U'\sin\theta' = \frac{1}{c}\Gamma(1 + \mathrm{B}\sin\theta')\, U'\sin\theta = \frac{U}{c}\sin\theta.$$

Thus, in any inertial frame of reference we can introduce the 4-vector

$$\vec{P}'\left(\frac{U'}{c}\cos\theta',\ \frac{U'}{c}\sin\theta',\ 0,\ i\frac{U'}{c}\right), \tag{7.31}$$

with $\vec{P}'^2 = 0$ in all reference frames.

It follows from the condition $\vec{P}'^2 = 0$ that the light wave *in vacuo* cannot be at rest in any inertial frame of reference. Comparing the components $\vec{P}'$ (7.31) to the components $\vec{k}'$ (7.8), we see that the transformation formulae for U' and ω' must be the same. This implies that the ratio U'/ω' must be invariant. Consequently, the energy of the same train of waves turns out to be different when measured by different observers. The ratio of energies is equal to the ratio of frequencies of the monochromatic radiation which forms the train of waves. The frequencies are determined by the same observers who measure the energy. The train is supposed to

be long enough since otherwise it will not be even approximately monochromatic.

From the formulae obtained the amplitude transformation law is easily established in the case of a plane wave. Indeed, from Eqs. (7.25) we get the following formula for energy density transformation:

$$w = w'\Gamma^2 (1 + \mathrm{B}\cos\theta')^2.$$

Comparing this expression to the frequency transformation formula (7.10)

$$\omega = \omega'\Gamma (1 + \mathrm{B}\cos\theta'),$$

we see that the energy density transforms as the square of frequency. Since the energy density is a quadratic function of field amplitudes of a plane wave, they transform according to the same rule as frequency does.

To illustrate the treatment of an electromagnetic wave as a system whose momentum and energy form a 4-vector, let us consider how the angular distribution of radiation from a dipole oscillator is transformed on transition from the frame K' in which the oscillator's centre of inertia is at rest to any other IFR. In the reference frame in which the oscillator's centre of inertia is at rest and the polar axis is directed along the oscillator's axis, the radiation intensity dI' in the direction (θ', φ') is known to be equal to $dI'(\theta', \varphi') = \text{const}\cdot\sin^2\theta'\, d\Omega'$. But the radiation intensity $dI = d\varepsilon/dt$, i.e. the energy radiated in a given direction per unit of time, is a relative quantity. Its transformation law is easy to establish; in the case of radiation $d\varepsilon = c\,dP$, where dP is a momentum fraction escaping with the radiation in a given direction, it being known that $d\varepsilon' = c\,dP'$. According to the Lorentz transformation

$$d\varepsilon' = \Gamma (d\varepsilon - \boldsymbol{V}\, d\boldsymbol{P}) = \Gamma (d\varepsilon - V\, dP\cos\theta) = \Gamma\, d\varepsilon (1 - \mathrm{B}\cos\theta),$$

$$dt' = \frac{1}{\Gamma}\, dt,$$

where dt' is the proper time. Having divided termwise the upper equality by the lower one, we get

$$dI = \frac{d\varepsilon}{dt} = \frac{d\varepsilon'}{dt'}\,\frac{1}{\Gamma^2 (1 - \mathrm{B}\cos\theta)} = \frac{dI'}{\Gamma^2 (1 - \mathrm{B}\cos\theta)}.$$

Then we immediately obtain the sought for result:

$$dI = \text{const}\cdot\frac{\sin^2\theta'\, d\Omega'}{\Gamma^2 (1 - \mathrm{B}\cos\theta)} = \text{const}\cdot\frac{\sin^2\theta'\, d\Omega}{\Gamma^4 (1 - \mathrm{B}\cos\theta)^3} =$$

$$= \text{const}\cdot\frac{\left(1 - \frac{V^2}{c^2}\right)^3 \sin^2\theta}{\left(1 - \frac{V}{c}\cos\theta\right)^5}\, d\Omega,$$

where we have used Eqs. (7.16) and (7.12′). It is seen from the obtained formula that the angular dependence of radiation in the frame K, relative to which the oscillator moves, differs essentially from the angular dependence in the frame K', especially in the case when $V \approx c$. In this case the maximum radiation is observed in a direction forming an acute angle with the oscillator's axis (Fig. 7.2).

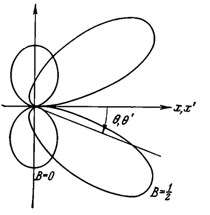


Fig. 7.2. The variation in the angular distribution of radiation emitted by a dipole oscillator on transition from the reference frame K' in which the oscillator is at rest ($B = 0$) to the reference frame K relative to which it moves ($B = 1/2$). The maximum radiation direction is seen to be inclined toward the oscillator's motion direction. The axis of the oscillator is oriented in the oscillator's motion direction.

It is worthwhile to consider how the radiation which is isotropic in the reference frame K' behaves in the frame K. All the necessary formulae are available now. In this case $dI'(\theta', \varphi') = \text{const}\cdot d\Omega'$ and

$$dI = \frac{dI'}{\Gamma^2(1 - B\cos\theta)} = \frac{\text{const}}{\Gamma^2(1 - B\cos\theta)}\,d\Omega' = \frac{\text{const}\cdot d\Omega}{\Gamma^4(1 - B\cos\theta)^2} = \text{const}\cdot\frac{\left(1 - \frac{V^2}{c^2}\right)^2}{\left(1 - \frac{V}{c}\cos\theta\right)^3}\,d\Omega.$$

From the last formula the "searchlight effect" in K can be seen. The radiation concentrates around the direction $\theta = 0$, since the value of the denominator is the least at $\cos\theta \approx 1$, with the ratio V/c fixed.

§ 7.4. The pressure exerted by an electromagnetic wave (light) on a surface. The pressure exerted on a surface of a body, i.e. the force acting on a unit of area, is defined by the momentum flowing across a unit of area and optical properties of the surface. The momentum flow is expressed via the spatial components of the energy-momentum-tension tensor $T_{\alpha\beta}$ which for a plane wave takes either the form of Eq. (7.17) or (7.19), depending on the propagation direction. When the wave propagates along the x' axis, then, as Eq. (7.17) shows, the tension tensor has only one component differing from zero, that is $T'_{xx} = -w'$. To find the momentum flow across the given surface element, one has to define the direction of the normal to this surface $\boldsymbol{n}(n_\alpha)$. Then (see Chapter 6) the momentum flow across the element dS with the normal $\boldsymbol{n}$

is equal to (Fig. 7.3)

$$T'_{\alpha\beta} n'_{\alpha} m'_{\beta}\, dS' = T'_{11} n'_{1} m'_{1}\, dS',$$

because only one term of the double summation differs from zero. The magnitude of pressure acting on the unit of area normal to the x' axis is equal to $p' = T'_{11} = T'_{xx} = |w'|$. If a light pulse propagates at the velocity c, the unit of area receives the energy per unit of time equal to $w'c = \mathscr{E}'$.

But we have seen that $w' = p'$, whence

$$p' = \mathscr{E}'/c. \qquad (7.32)$$

Consequently, the pressure of light is equal to the energy of an electromagnetic wave incident on a unit of area per unit of time and divided by c provided that the wave is absorbed.

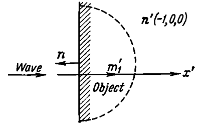


Fig. 7.3. The calculation of pressure exerted by an electromagnetic wave on a surface.

Now let us determine the force exerted on a wall by a light wave incident on this wall at a certain angle and reflected from it. Let the incident angle be equal to θ. We shall denote the normal to the wall by $\boldsymbol{n}$, and the unit vectors directed along the propagation of the incident and reflected waves by $\boldsymbol{s}$ and $\boldsymbol{s}'$ respectively. The momentum flow across a unit of area will yield the pressure $\boldsymbol{p}$ whose components are as follows:

$$p_{\beta} = T_{\alpha\beta} n_{\alpha} + T^{*}_{\alpha\beta} n_{\alpha} = (T_{\alpha\beta} + T^{*}_{\alpha\beta})\, n_{\alpha},$$

where $T_{\alpha\beta}$ and $T^{*}_{\alpha\beta}$ are the tension tensor components of incident and reflected waves.

The components of $T_{\alpha\beta}$ for the wave propagating at the angle θ to the x axis are given by Eq. (7.19). The three-dimensional wave vector of the reflected beam differs from that of the incident beam by the substitution of $-\theta$ for θ. Let us introduce the reflection coefficient R so that $w^{*} = Rw$. Keeping in mind that $T_{11} = T'_{44} \cos^2\theta = -w\cos^2\theta$, and $T_{12} = T'_{44} \sin\theta\cos\theta = -w\sin\theta\cos\theta$, we obtain the following expression for the normal force (pressure of light):

$$p_x = (w + Rw)\cos^2\theta = w(1 + R)\cos^2\theta$$

and for the tangential force:

$$p_y = w(1 - R)\sin\theta\cos\theta.$$

Let us write out the magnitudes of normal pressure p_x for the two most interesting cases. In the case of the normal incidence ($\theta = 0$) the pressure is equal to $2w$ if the wave is reflected completely and to w if it is fully absorbed. In the case of isotropic radiation p_x should be averaged over all directions, i.e. the average value of $\cos^2\theta$ should be taken. But the average value of the square of the direction cosine of the spatially-isotropic unit vector is equal to $^1/_3$. Thus in the case of a total absorption of isotropic radiation the pressure is defined by the formula $p = w/3$.

Surely, all these formulae can be obtained by means of elementary reasoning. Proceeding from the magnitude of electromagnetic field momentum density $g = S/c^2$ and that of the Poynting vector in a plane wave $S = wc$, we obtain $g = w/c$ (w is the energy density). When a plane wave falls on the wall at the angle θ, a unit of area takes up all the energy and momentum per unit of time which are confined within an oblique cylinder whose base is formed by the unit of area and whose generatrix is numerically equal to the light propagation velocity. The volume of such a cylinder is equal to $c\cos\theta$. Therefore 1 m^2 of the wall takes up the energy $\mathscr{E} = wc\cos\theta$ during 1 s while the momentum G_x transmitted in the direction perpendicular to the wall during 1 s is equal to $G_x = g\cos\theta \cdot c\cos\theta = w\cos^2\theta$. But the momentum transmitted to a unit of area per unit of time is just equal to the pressure p. Introducing as before the reflection coefficient R, we get $p = w(1 + R)\cos^2\theta$.

§ 7.5. The light frequency variation on reflection from a moving surface (mirror). Let in the frame K a light beam propagate at the angle θ_0 to the x axis in the (x, y) plane. A mirror located parallel to the y axis moves relative to the reference frame K at the velocity V. The light beam reaching the mirror is reflected from it. We shall find the frequency and propagation direction of the reflected beam in terms of the frame K.

It is convenient to introduce the reference frame K' fixed to the mirror. Then the problem is solved as follows. In the frame K the 4-vector of the light beam is specified, i.e. the frequency and propagation direction of light are known. The frequency of light and the beam direction in the frame K' are easy to find using the Lorentz transformation formulae. In the frame K' in which the mirror is at rest the routine law of reflection is valid: an angle of incidence is equal to an angle of reflection. This implies that the 4-vector of the reflected beam differs from that of the incident beam only by the sign of the wave vector component along the x axis. To obtain the 4-vector of the reflected beam in the frame K, one has to apply the Lorentz transformation once more.

Now, let in the frame K the light beam of the frequency ω_0 propagate at the angle θ_0 to the x axis in the plane (x, y). The

components of the 4-vector $\vec{k}_0$ in the frame K will be

$$k_1^0 = k^0 \cos\theta_0 \equiv \frac{\omega_0}{c}\cos\theta_0, \quad k_3^0 = 0,$$
$$k_2^0 = k^0 \sin\theta_0 \equiv \frac{\omega_0}{c}\sin\theta_0, \quad k_4^0 = i\frac{\omega_0}{c} = ik^0. \tag{7.33}$$

Let us find the 4-vector $\vec{k}'$ of the same beam in the frame K'. According to the Lorentz transformation (4.10a) we get

$$k_1' = \Gamma(k_1^0 + iBk_4^0), \quad k_2' = k_2^0, \quad k_3' = k_3^0, \quad k_4' = \Gamma(k_4^0 - iBk_1^0). \tag{7.34}$$

The component k_1' alters the sign on reflection from the mirror which is stationary in the frame K'. Therefore, the 4-vector $\vec{k}''$ of the reflected beam will take the form

$$k_1'' = -\Gamma(k_1^0 + iBk_4^0), \quad k_2'' = k_2^0, \quad k_3'' = k_3^0, \quad k_4'' = \Gamma(k_4^0 - iBk_1^0). \tag{7.35}$$

In the frame K the reflected beam will be described by the 4-vector $\vec{k}$ which is derived from the 4-vector $\vec{k}''$ via the inverse Lorentz transformation from the frame K' to the frame K:

$$k_1 = \Gamma(k_1'' - iBk_4'') = -\Gamma^2\{(1+B^2)k_1^0 + 2iBk_4^0\} =$$
$$= -\Gamma^2\frac{\omega_0}{c}\{(1+B^2)\cos\theta_0 - 2B\}, \tag{7.36}$$

$$k_2 = k_2^0 = \frac{\omega_0}{c}\sin\theta_0, \quad k_3 = k_3^0 = 0, \tag{7.37}$$

$$k_4 = \Gamma(k_4'' + iBk_1'') = \Gamma^2\{(1+B^2)k_4^0 - 2iBk_1^0\} =$$
$$= i\Gamma^2\frac{\omega_0}{c}\{(1+B^2) - 2B\cos\theta_0\}. \tag{7.38}$$

Since $k_3 = k_3^0 = 0$, the reflected beam keeps propagating within the plane (x, y). Assuming that

$$k_1 = \frac{\omega}{c}\cos\theta, \quad k_2 = \frac{\omega}{c}\sin\theta,$$
$$k_3 = 0, \quad k_4 = i\frac{\omega}{c}, \tag{7.39}$$

we get from Eq. (7.38)

$$\frac{\omega}{\omega_0} = \frac{(1+B^2) - 2B\cos\theta_0}{1-B^2}. \tag{7.40}$$

Consequently, the frequency ω of reflected light observed in the frame K is not equal to the frequency ω_0 of incident light.

In the frame K the tangent of reflection angle is obtained in the following form:

$$\tan\theta = \frac{k_2}{k_1} = -\frac{\sin\theta_0 (1 - B^2)}{(1 + B^2)\cos\theta_0 - 2B}. \tag{7.41}$$

It is seen from Eq. (7.41) that $\theta \neq \theta_0$. Therefore, the angle of incidence and the angle of reflection prove to be different in the frame K (Fig. 7.4*a*).

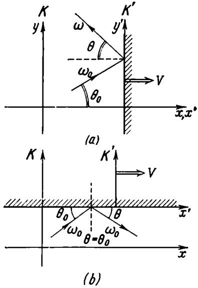


Fig. 7.4. Reflection of light from a moving mirror (*a*) A mirror moves along the normal to its plane As a result of reflection, the frequency changes and the angle of incidence is not equal to the angle of reflection. (*b*) A mirror moves parallel to its plane The frequency remains unchanged on reflection; the angle of incidence is equal to the angle of reflection

It is worthwhile to write out the formulae pertaining to the normal incidence of light on the mirror. Let in the frame K the angle of incidence $\theta_0 = 0$. Then we get

$$\omega = \omega_0 \frac{1 - B}{1 + B}, \quad \cos\theta = -1,$$
$$\sin\theta = 0. \tag{7.42}$$

It follows from Eq. (7.42) that reflected light also propagates along the normal to the mirror, although in the direction opposite to the initial one. The frame K' in which the mirror is at rest moved in the same direction in which light propagated. The frequency of light decreased on reflection.

If the mirror moves toward the beam, the quantity B alters its sign, and therefore we get

$$\omega = \omega_0 \frac{1 + B}{1 - B}, \quad \cos\theta = -1,$$
$$\sin\theta = 0.$$

The frequency of light increases on reflection.

Making use of this effect, one can determine the velocity of a moving object, e.g. an automobile. When an automobile moves toward an observer, the frequency change on reflection is found from Eq. (7.40) with an accuracy to within the terms of the order of B^2 ($\Delta\omega = \omega - \omega_0$):

$$\frac{\Delta\omega}{\omega_0} = \frac{2B}{1 - B} \approx 2B. \tag{7.43}$$

If, for example, the velocity of an automobile is 72 km/h = 20 m/s, then $B = 20/3 \cdot 10^8 \approx 0.6 \cdot 10^{-7}$. This relative change of frequency can readily be detected by means of standard instruments.

We have considered the case when the mirror moves along its normal. However, the mirror may move parallel to its own plane (Fig. 7.4*b*). In this case we have to modify somewhat the formulae used. Eqs. (7.33) and (7.34) remain unchanged; however, the sign of k_2 is altered on reflection, and therefore,

$$k_1'' = \Gamma\left(k_1^0 + iBk_4^0\right), \quad k_2'' = -k_2^0, \quad k_3'' = k_3^0, \quad k_4'' = \Gamma\left(k_4^0 - iBk_1^0\right). \tag{7.44}$$

Since $k_3 = k_3^0 = 0$, the reflected beam remains in the plane (x, y) as before. Returning to the frame K, we get

$$k_1 = \Gamma\left(k_1'' - iBk_4''\right) = k_1^0, \tag{7.45}$$

$$k_2 = -k_2^0, \tag{7.46}$$

$$k_3 = 0, \tag{7.47}$$

$$k_4 = \Gamma\left(k_4'' + iBk_1''\right) = k_4^0. \tag{7.48}$$

From Eq. (7.48) we immediately obtain that $\omega = \omega_0$, and while writing $\tan\theta = k_2/k_1$, we find that $\tan\theta = -\tan\theta_0$, i.e. $\theta = -\theta_0$. Hence, when the mirror moves parallel to itself the frequency of incident light is equal to that of reflected light, and the angle of incidence is equal to the angle of reflection (in the frame K).

In conclusion let us write out the formulae describing the reflection from the mirror moving along the normal to its plane in the non-relativistic approximation, that is in the case when the velocity of the mirror is small: $B = V/c \ll 1$. Ignoring all terms of the order B^2, we get

$$\omega = \omega_0(1 - 2B\cos\theta_0),$$
$$\cos\theta = -\cos\theta_0 + 2B\sin^2\theta_0,$$
$$\sin\theta = \sin\theta_0 + 2B\sin\theta_0\cos\theta_0.$$

In the case of the normal incidence on the mirror ($\theta_0 = 0$) the beam is reflected in the direction opposite to the initial one whereas the frequency varies according to the following law:

$$\omega = \omega_0\left(1 - 2\frac{V}{c}\right), \tag{7.49}$$

if the mirror moves in the same direction as the light beam. If the mirror moves toward light, then

$$\omega = \omega_0\left(1 + 2\frac{V}{c}\right). \tag{7.50}$$

Eqs. (7.49) and (7.50) permit of a straightforward interpretation. Reflected light can be imagined as going from an imaginary source positioned behind the mirror, with the velocity of this imaginary source being equal to $2V$. Consequently, if the imaginary source is replaced by the real one with the same natural frequency ω_0, the frequency change defined by Eqs. (7.49) and (7.50) will correspond to the Doppler effect for this source.

The considered instances of light reflected from a moving mirror represent specific cases of the general problem involving electromagnetic phenomena arising at a moving interface dividing two media.

§ 7.6. Light quanta (photons) as relativistic particles. Relativistic mechanics presented in Chapter 5 dealt with particles possessing a finite (differing from zero) rest mass. This is manifested, in particular, by a 4-momentum of a particle $\vec{P} = m\vec{V}$ being meaningful only on the condition that $m \neq 0$. The particles whose rest mass differs from zero are referred to as *tardyons*. All such particles cannot reach the velocity c through acceleration. This can be seen from the fact that the infinite energy and momentum are required for these particles to gain the ultimate velocity ($\mathscr{E} = mc^2\gamma$, $\boldsymbol{p} = m\gamma\boldsymbol{v}$, but if $v \to c$, the factor $\gamma \to \infty$). The solutions of all concrete problems cited in Chapter 5 testify that v remains less than c in all cases.

While investigating an electromagnetic field interacting with microparticles, physicists inferred that in such an interaction a microparticle, e.g. an electron, always gains a definite energy and definite momentum from the electromagnetic field. (For the sake of simplicity we discuss here a monochromatic radiation, i.e. a radiation of a given frequency ω.) For the first time the assumption about an electromagnetic field imparting energy to an electron by definite portions (quanta) was made by A. Einstein in the framework of the photoeffect theory (1905). In order to explain the scattering of high-energy γ-quanta by electrons, an electromagnetic field had to be assumed to transfer not only a definite energy to electron but a definite momentum as well (the Compton effect, 1923).

These properties of an electromagnetic field interacting with an electron can be graphically described in terms of an interaction of "particles of light", possessing a definite energy and momentum, with an electron. Of course, it would be extremely naive to imagine an electromagnetic field consisting of some kind of particles resembling billiard balls. This "particle-of-light" concept is perfectly suitable for describing energy and momentum exchange between a field and microparticles. This being borne in mind, the concept of particles of light, called *light quanta*, or *photons*, cannot lead to a misunderstanding.

What properties should we attribute to a photon in order to treat it as a relativistic particle? One of the photon's properties, to wit, the relationship between its energy and momentum, can be obtained from macroscopic electrodynamics. Indeed, the pressure P of light falling on a wall from vacuum was given by the relation (7.32)

$$P = \mathscr{E}/c, \tag{7.51}$$

where $\mathscr{E}$ denotes the energy taken up by a unit of area of the wall per unit of time. Now let us imagine that photons fall on the wall. (Below it will be seen that the most essential is the fact that the energy and momentum are transmitted by quantified portions.) Let us depict light as a plane wave so that all photons move in the same direction. Suppose each photon carries the energy ε and momentum p. If a unit of area of the wall absorbs N photons falling on it every unit of time, the wall gains the energy $N\varepsilon$ and momentum Np. But the momentum gained by a unit of area of the wall per unit of time is just the pressure of light P* so that $P = Np$ and $\mathscr{E} = N\varepsilon$. That is why from Eq. (7.51) follows the relation between the energy and momentum of a photon:

$$p = \varepsilon/c. \tag{7.52}$$

But in the case of relativistic particles the relationship between the energy, momentum and velocity of motion is established by the expression $\boldsymbol{p} = (\varepsilon/c^2)\boldsymbol{v}$, whence it is clear that the relation (7.52) is valid only if $v = c$. Thus, a photon can be interpreted as a relativistic particle provided it moves at the velocity c.

Just as for any relativistic particle, the 4-vector of energy-momentum $\vec{P}(\boldsymbol{p}, i\varepsilon/c)$ can be constructed for a photon *in vacuo*. Using the general formulae for the calculation of the square of the 4-vector and taking into account Eq. (7.52), we get $\vec{P}^2 = 0$. On the other hand, for conventional particles $\vec{P}^2 = -m^2c^2$ (see Eq. (5.47)). This means that the rest mass of a photon is equal to zero. In order to permit the (imaginary) particle to reach the ultimate relativistic velocity, we had to discard the finite rest mass.

The rest mass of a photon proved to be equal to zero, and at first sight this fact seems to be rather regrettable. We have got used to all bodies and particles in nature to possess a rest mass. Until quite recently mass was regarded as an indispensable attribute of matter taken for actually existing reality. Physicists were also inclined to believe that a rest mass defines individual features of every body or object. In classical mechanics any mass-possess-

* According to Newton's law $\boldsymbol{F} = d\boldsymbol{p}/dt$. Dividing both sides of this equality by the area on which the force acts, we obtain the pressure, i.e. $p = F/s = (1/s)\,(dp/dt)$; it is the momentum increment transmitted to a unit of area per unit of time that appears on the right-hand side here.

ing entity could be traced, at least theoretically, in the course of time.

Until the beginning of this century light was thought of as a mysterious phenomenon; even physicists doubted its material nature. But in 1901 P. N. Lebedev discovered the pressure of light experimentally. The pressure is caused by the momentum flux. Prior to this, there were no particular doubts that light carried energy. But if light possesses both energy and momentum, its material nature cannot be questioned. Although the rest mass of an individual light quantum is equal to zero, there is nothing reprehensible about it. In nature there are objects whose rest mass is finite, and also objects of zero rest mass. The latter move at the velocity of light and cannot be stopped; throughout all reference frames their velocity is the same. When brought to a standstill, they terminate their existence, passing into other forms of matter. It is the very fact that the forms of matter possessing a zero rest mass convert into those with a finite rest mass (and back), that illustrates an equivalence of these forms.

The expressions for the photon's energy and momentum cannot be obtained in terms of relativistic mechanics. Contemporary physics, however, discovered that in the processes of emission and absorption, as well as in interactions with matter, light behaves as an assembly of quasiparticles, each of which possesses the energy $\hbar\omega$ and momentum $\hbar\omega/c$. Here $\hbar$ is Planck's constant, $\hbar = 6.626\cdot 10^{-34}$ J·s, and ω is the circular frequency of light. Every elementary act of interaction with matter involves one quasiparticle of this kind, called a light quantum by Einstein in his time. The energy and momentum conservation laws are valid in these interactions. The expression $\varepsilon = \hbar\omega$ can be taken for the light quantum energy and $\boldsymbol{p} = \frac{\hbar\omega}{c}\boldsymbol{s}$ for its momentum; here $\boldsymbol{s}$ is a unit vector directed along the light beam. Thus, if one treats a light quantum (photon) as a relativistic particle, its 4-vector energy-momentum takes the form $\vec{P}\left(\hbar\boldsymbol{k}, i\frac{\hbar\omega}{c}\right)$, where $\boldsymbol{k} = k\boldsymbol{s}$, $k = 2\pi/\lambda = \omega/c$. Cancelling all components of $\vec{P}$ by the common factor, which is Planck's constant $\hbar$, we obtain the same 4-vector $\vec{k}$ again which earlier denoted the wave vector. This time, however, it is defined to fit a photon: $\vec{k}\,(\boldsymbol{k}, ik)$, $k = \omega/c$, $\vec{P} = \hbar\vec{k}$. Since the four-dimensional momentum of a photon coincides, with an accuracy to within the factor $\hbar$, with the four-dimensional wave vector introduced in § 7.2, all results obtained for a wave are fully applicable to a photon. We mean here the formulae describing the Doppler effect, aberration of light, change of light frequency on reflection from a moving mirror. In terms of the photon theory of

light one can easily derive the formula describing the pressure of light. Indeed, let a photon fall on a surface of a body at an angle θ. The normal component of its momentum is equal to $\hbar\omega \cos\theta/c$ (Fig. 7.5). On absorption of the photon the wall acquires just this momentum along the normal's direction. If the photon is reflected, the magnitude of the transmitted momentum depends on the reflection coefficient, that is the photon reflection probability; let us denote it by R. Then the momentum component transmitted normally to the wall on reflection of the photon is equal to $(1+R)(\hbar\omega/c)\cos\theta$ where $R \leqslant 1$. If n denotes the number of photons in 1 m^2, all the photons confined within an oblique cylinder whose generatrix is equal to c will fall on 1 m^2 area of the wall per 1 s. The base of this cylinder is equal to 1 m^2 and its volume to $c\cos\theta$. Consequently, $nc\cos\theta$ photons will fall on 1 m^2 area of the wall per 1 s. Provided all these photons get absorbed, the wall acquires the energy $\hbar\omega nc\cos\theta$ and the normal momentum component

$$nc\cos\theta(1+R)\frac{\hbar\omega}{c}\cos\theta = \hbar\omega n(1+R)\cos^2\theta.$$

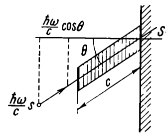


Fig. 7.5. Calculation of pressure produced by light The area of the oblique cylinder base is equal to $S = 1$.

The momentum transmitted to a unit of area of the wall per unit of time is just the pressure on the wall; therefore, $p = w(1+R)\cos^2\theta$ where $w = n\hbar\omega$ is the energy density in the beam. This result coincides with the one derived in § 7.4.

Unfortunately, today a "photon mass" is still defined by the formula taken from relativistic mechanics of particles with a finite rest mass, namely, $m_{ph} = \varepsilon/c^2$ $(\varepsilon = \hbar\omega)$. First of all, the formula $\varepsilon = mc^2$ holds explicitly for the case $m \neq 0$ (see Chapter 5) and is quite irrelevant to the case $m = 0$. Besides, the mass m_{ph} has no physical sense at all. There is also no sense in speaking of the inertial properties of a photon: in all reference frames it moves at the velocity c, i.e. a photon *in vacuo* cannot be either accelerated or decelerated (it can only be exterminated). In quantum statistics photons are treated as identical particles, this assumption giving correct results. However, if a "photon mass" had any meaning, the "blue" photon would have been "heavier" than the "red" one, thus violating the identity of particles. On the contrary, the only common property of photons is their zero rest mass. Sometimes the photon mass m_{ph} is used to explain the deviation of a light beam in a gravitational field. It is inconsistent. however, to treat such

a relativistic object as light in terms of classical mechanics. Needless to say, the relativistic theory predicts the deviation of a light beam in a gravitational field without resorting to any photon mass. Finally, there are some who wish to have the "mass conservation law" in relativistic physics. Since a rest mass is not additive (§ 5.6), the new "masses" are introduced on the basis of the relation $m = \mathscr{E}/c^2$. This is, however, quite meaningless to undertake since the conservation of such a "mass" is just a consequence of the energy conservation law which is always valid. To summarize, we can state that the introduction of a photon mass does not bring any advantages, complicating needlessly the subtle concept of mass (see §§ 5.6 and 5.7).

Coming back to the zero rest mass of a photon, let us make some more remarks. There is no real IFR in which a photon would be at rest, so that the photon's rest mass is an unobserved quantity. Just as meaningless is to speak of time flowing in the reference frame fixed to a photon. The zero mass of a photon does not at all signifies the absence of mass. For example, the temperature 0 °C does not mean that the body lacks internal energy. It should be recalled here that in the STR there are the world lines of the zero length which are not less meaningful than all other lines. Surely, this is all due to the velocity of light being distinguished among other velocities. Apart from photons *in vacuo* there are also the "real" particles, neutrinos, moving at the velocity c as well. Their rest mass is also equal to zero and cannot be observed in an experiment. After all, the question whether the photon's mass is equal to zero or not can be solved experimentally. There are methods capable of detecting the photon's rest mass if it differs from zero. As more and more experiments of this kind are conducted, the lower limit of the photon's "rest mass" slides gradually lower and lower. By the end of 1975 this limit reached the value $m < 10^{-63}$ kg.

§ 7.7. Light quanta in a medium. The Vavilov-Cherenkov effect. The anomalous Doppler effect. It is seen from the previous section and § 6.12 that the photon's momentum in a medium is determined according to Eq. (6.183) when proceeding from the Minkowski tensor and according to Eq. (6.184) when proceeding from the Abraham tensor. The photon's energy remains constant on transition from one medium to another, provided the oscillation frequency does not vary. Then what expression for the momentum should be used when the momentum conservation law is applied to "light quanta in a medium"? There is no direct answer to this question, and some considerations in this respect will be given at the end of this section. And now we shall show that if light quanta (photons) in a medium are utilized in the form given by Eq. (6.183), we can obtain useful results concerning radiation kinematics, i.e. conditions imposed on a frequency and direction of ra-

diation. These conditions are defined by the energy and momentum conservation laws. We shall begin with the elementary derivation of the conditions for the Vavilov-Cherenkov radiation.

In this case radiation is emitted by a particle which has no internal degrees of freedom. We shall write down the conservation laws for an electron-radiation system. Of course, the conservation laws *per se* do not answer the question as to whether radiation will occur. This question can be solved by calculation on the basis of equations of electrodynamics. However, if the conservation laws do not hold, the radiation is absent *a fortiori*.

Suppose a light quantum has been radiated. If the energy and momentum of an electron before the radiation were $\mathscr{E}_0$, $\boldsymbol{p}_0$ and became $\mathscr{E}_1$, $\boldsymbol{p}_1$ after it, the energy and momentum conservation laws take the form

$$\Delta\mathscr{E} = \mathscr{E}_0 - \mathscr{E}_1 = \hbar\omega, \tag{7.53}$$

$$\Delta\boldsymbol{p} = \boldsymbol{p}_0 - \boldsymbol{p}_1 = \frac{\hbar\omega}{c} n\boldsymbol{s}. \tag{7.54}$$

Written in such a form, the conservation laws presuppose that the change in the energy and momentum of an electron is connected only with radiation. We can easily find the required consequence of Eqs. (7.53) and (7.54) recalling that according to Newton's law $\Delta\boldsymbol{p} = \boldsymbol{F}\cdot\Delta t$; multiplying both sides of this relation by $\boldsymbol{v}$ and recalling that $\boldsymbol{F}\boldsymbol{v}\,\Delta t = \Delta\mathscr{E}$, we get

$$\Delta\mathscr{E} = \boldsymbol{v}\,\Delta\boldsymbol{p}. \tag{7.55}$$

Naturally, this relation is valid only for small momentum changes.

Substituting Eqs. (7.53) and (7.54) into Eq. (7.55), we can cancel out $\hbar\omega$, then note that $\boldsymbol{v}\boldsymbol{s} = v\cos\theta$ where θ is the angle between the directions of the electron motion and the radiation propagation. Then the final kinematic condition for the radiation angle θ will be written as follows:

$$\cos\theta = \frac{c}{nv}. \tag{7.56}$$

This is the condition for the Vavilov-Cherenkov radiation. It is not satisfied if an electron moves uniformly *in vacuo* ($n = 1$), since $|\cos\theta| \leqslant 1$, and the electron's velocity v is always less than c. Consequently, an electron moving uniformly *in vacuo* does not radiate.

We obtained this result directly from the principle of relativity: a charge resting in a certain IFR does not radiate. This charge moves uniformly and rectilinearly relative to any other IFR. Radiation, however, either occurs in all IFRs or does not occur in any of them. Consequently, a uniformly moving electron does not radiate. This reasoning is not correct, however, for an electron moving in a medium since a new characteristic velocity, that is

the velocity of an electron relative to a medium, appears here, defining at the same time the "privileged" reference frame fixed to that medium.

Note that in our approximation the final results do not contain the quantity $\hbar$ in spite of the quantum mechanical ideas used. The result obtained is a classical one. The quantum mechanical ideas were employed for purely methodical reasons. We make use of the two conservation laws which do not require an obligatory utilization of quantum mechanical concepts.

It is easy to obtain the radiation condition with the recoil momentum taken into account. Let us introduce the 4-vector of energy-momentum (or briefly, 4-momentum) of a light quantum in a medium

$$\vec{\pi}\left(\frac{\hbar\omega}{c} n\mathbf{s},\ i\frac{\hbar\omega}{c}\right). \tag{7.57}$$

The photon radiation by an electron must obey the conservation law for the 4-vector of energy-momentum, or, in other words, the energy and momentum conservation laws. Let the 4-momentum of an electron before the radiation be equal to $\vec{p}_0$, after the radiation $\vec{p}$, while the 4-momentum of the light quantum to $\vec{\pi}$, that is

$$\vec{p}_0\,(m\gamma_0\mathbf{v}_0,\ im\gamma_0 c),\quad \vec{p}\,(m\gamma\mathbf{v},\ im\gamma c),\quad \vec{\pi}\left(\frac{\hbar\omega}{c} n\mathbf{s},\ i\frac{\hbar\omega}{c}\right).$$

The conservation law for the 4-momentum has the form

$$\vec{p}_0 = \vec{p} + \vec{\pi},$$

or in components,

$$p_{0i} - \pi_i = p_i.$$

Squaring the latter relation, we get

$$p_{0i}^2 - 2p_{0i}\pi_i + \pi_i^2 = p_i^2,$$

where each term involves the summation over the index i. However, due to the invariance of the particle momentum square $p_{0i}^2 = p_i^2$, and we get

$$\pi_i^2 = 2p_{0i}\pi_i. \tag{7.58}$$

Cancelling by π_i is prohibited here, of course: the left-hand and right-hand sides involve independent summation. We shall calculate the left-hand and right-hand sides of Eq. (7.58) separately:

$$p_{0i}\pi_i = \frac{\hbar\omega}{c} nm\gamma_0(\mathbf{v}_0\mathbf{s}) - m\gamma_0\hbar\omega,\quad \pi_i^2 = \left(\frac{\hbar\omega}{c}\right)^2(n^2-1).$$

Equating these expressions and taking into account that $\mathbf{v}_0\mathbf{s} = v_0\cos\theta$, where θ is the angle between emitted light and the

electron propagation direction, we obtain

$$\frac{1}{2}\left(\frac{\hbar\omega}{c}\right)^2(n^2-1)=\frac{\hbar\omega}{c}nm\gamma_0 v_0\cos\theta - m\gamma_0\hbar\omega.$$

Hence

$$\cos\theta=\frac{1}{nm\gamma_0 v_0}\left\{\frac{\hbar\omega}{2c}(n^2-1)+m\gamma_0 c\right\}=$$
$$=\frac{c}{nv_0}\left\{1+\frac{\hbar\omega}{2mc^2}(n^2-1)\sqrt{1-\frac{v_0^2}{c^2}}\right\}.$$

When considering absorption of a quantum instead of its radiation, we should alter the sign in front of $\hbar\omega$ in the latter formula. When $\hbar\omega/mc^2 \ll 1$, which is true for visible light and an electron, we get back to the classical condition for radiation (Eq. 7.56):

$$\cos\theta=\frac{c}{nv_0}=\frac{V_{ph}}{v_0}.$$

Radiation kinematics deals also with the problem of light changing its frequency and propagation direction on transition from one IFR to another. We mean the Doppler effect and aberration here. Surely, these problems are solved easier in terms of the STR. In § 7.2 we considered the propagation of light *in vacuo*. Here we shall obtain the requisite formulae for a uniform and isotropic medium whose refraction index is equal to n; the formulae to be obtained will prove to be quite different from the case of vacuum.

In fact, all calculations differ only slightly from those performed in § 7.2 and therefore we shall present them only in a brief form; in return, we shall discuss the obtained results in detail. From the 4-vector $\vec{\pi}$ we obtain the 4-vector of a photon in a medium, proportional to the former vector:

$$\vec{k}\left(\frac{\omega}{c}ns,\quad i\frac{\omega}{c}\right),\quad \vec{\pi}=\hbar\vec{k}. \tag{7.59}$$

Assume that in the reference frame K' light propagates in the plane (x', y') at the angle θ' to the x' axis; a medium is at rest in the frame K'. Then

$$\vec{k}'\left(\frac{\omega'}{c}n\cos\theta',\quad \frac{\omega'}{c}n\sin\theta',\quad 0,\quad i\frac{\omega'}{c}\right). \tag{7.60}$$

The components of the vector $\vec{k}$ in the frame K are to be found from the same Eqs. (7.9) from which it is seen that the beam remains, as before, in the plane (x, y) of the frame K. Instead of Eq. (7.10) we shall get

$$\omega = \omega'\Gamma(1+\mathrm{B}n\cos\theta'), \tag{7.61}$$

and instead of Eqs. (7.11) and (7.12)

$$n\cos\theta = \frac{n\cos\theta' + B}{1 + Bn\cos\theta'}, \tag{7.62}$$

$$\sin\theta = \frac{\sin\theta'}{\Gamma(1 + Bn\cos\theta')}. \tag{7.63}$$

Hence, the final formula for the Doppler effect takes the form (cf. Eq. (7.13))

$$\omega = \frac{\omega'\sqrt{1 - B^2}}{1 - Bn\cos\theta} \tag{7.64}$$

as before, $\omega' = \Gamma(\omega - \boldsymbol{kV})$ (cf. Eq. (7.14)), and for the aberration angle (cf. Eq. (7.15))

$$\tan\theta = \frac{n\sqrt{1 - B^2}}{B + n\cos\theta'}\sin\theta'. \tag{7.65}$$

Let a monochromatic source possessing the natural frequency ω_0 be at rest in the frame K', i.e. move uniformly at the velocity V

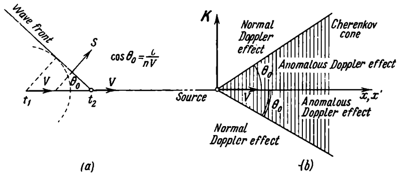


Fig. 7.6. (a) The Vavilov-Cherenkov radiation and the kinematic explanation of its generation The positions of a uniformly moving particle are shown for the moments of time t_1 and t_2. In the time interval $t_2 - t_1$ the wave front will take the position shown by a dotted circumference line (b) The Cherenkov cone divides the space around the source into the regions of the anomalous and normal Doppler effect

relative to K. Then $\omega' = \omega_0$. It is seen from Eq. (7.64) that for $n > 1$, i.e. for such common media as, for example, water and glass, $Bn\cos\theta$ may exceed unity even at $V < c$, and therefore the denominator may turn into zero or even become a negative quantity. Since an altered frequency sign implies, at the most, only an alteration of an oscillation phase:

$$\cos(-\omega t) = \cos\omega t, \quad \sin(-\omega t) = -\sin\omega t = \sin(\omega t + \pi/2),$$

a frequency may be always regarded as a positive quantity. Consequently, the formula for the Doppler effect in a medium can be finally written down as follows:

$$\omega = \frac{\omega_0 \sqrt{1 - B^2}}{|1 - Bn \cos\theta|}. \tag{7.66}$$

First of all, note that from Eq. (7.66) it is seen that a medium does not affect the transverse Doppler effect: in the case of $\theta = \pi/2$ we get exactly the same Eq. (7.14) as for vacuum. Thus, we get one more evidence that the transverse Doppler effect arises only due to the relativity of time intervals between events.

When $1 - Bn\cos\theta = 0$, the denominator of Eq. (7.66) turns into zero, this being the condition for the Cherenkov radiation. A moving charged particle having no internal degrees of freedom radiates within a cone around its propagation direction. In the case of a neutral source the Cherenkov cone divides space into two parts with respect to the observed Doppler effect. The condition $1 - Bn\cos\theta > 0$ is valid outside the Cherenkov cone (Fig. 7.6) where we observe the normal Doppler effect for which $(d\omega/d\theta) > 0$, as it is always the case *in vacuo*. "Within" the Cherenkov cone $1 - Bn\cos\theta < 0$ and $(d\omega/d\theta) < 0$; this corresponds to the anomalous Doppler effect.

It is interesting to compare the formulae for aberration of light *in vacuo* and in a medium. In the case of a normal incidence in the frame K' the aberration angle α in the frame K is determined from the following relation:

$$\tan\alpha = \frac{B}{\sqrt{1 - B^2}} \frac{1}{n}. \tag{7.67}$$

The only difference from the case of vacuum (cf. Eq. (7.15)) consists in the refraction index n appearing in the denominator. There is no singularity at $V = c/n$.

And in conclusion we shall decide which expression for the photon's momentum should be regarded as "correct". As it was pointed out in § 6.12, two different expressions for the photon's momentum in a medium, $p^M = \frac{h\omega}{c} n$ and $p^A = \frac{h\omega}{cn}$, correspond to different subdivisions of the momentum density of an electromagnetic field in a medium into "the momentum density of a field" and "the momentum density of a medium itself". Since considering the Cherenkov effect we are interested in a complete momentum being lost by an electron, and such a momentum is defined by the expression g^M, the employment of g^M leads to the correct result *.

* For more details see V. L. Ginzburg, UFN **110**, 309 (1973); V. L. Ginzburg, V. A. Ugarov, UFN **118**, 175 (1976).

CHAPTER 8

ON CERTAIN PARADOXES OF THE SPECIAL THEORY OF RELATIVITY

If one consults a dictionary, he learns that the word "paradox" has at least three meanings: an unusual judgement differing from a generally accepted one by its originality; a surprising deduction from certain assumptions; and a result which seems incredible at first glance but proves to be correct on more careful consideration. Studying the STR, we can easily find examples illustrating all meanings of the word "paradox".

As to the "generally accepted opinion", it is the ideas of classical physics as a whole and the classical concepts of properties of space and time specifically, that are applied to the STR. The classical ideas of space and time coincide to a great extent with those concepts that we acquire in our schoolyears and in our everyday practical life. These customary ideas have long become generally accepted, and their employment rests, we believe, on the "common sense". But eventually, the common sense is an accumulation of our settled down convictions among which there may also be false ones. These false convictions are being exposed as science develops. Very often it turns out that certain ideas are true only approximately, and the field of their application proves to be limited. That is exactly what happened to some concepts when the STR appeared.

In everyday life and in classical mechanics, for example, we got accustomed to time having absolute meaning. Of course, this is substantiated fairly well. The theory of relativity showed that the time moments of an event measured in different IFRs are different, i.e. relative. However, it is psychologically difficult to readjust oneself to relativity of time, especially as this relativity manifests itself only at relativistic velocities (which macroscopic bodies never attain) and never in everyday life. Relativity of time, relativity of simultaneity and time intervals between events correlate with relativity of lengths of scales moving with respect to one another From the viewpoint of the common sense which makes us believe in absolute time, these conclusions are paradoxical. In terms of contemporary physics they are not paradoxes at all. Relativity of time is just a modern interpretation of measurement results ob-

tained for the time of an event. By the way, this representation conforms very well to the treatment of time in terms of dialectical materialism according to which time, as a form of existence of eternally moving matter, may depend on the motion of matter.

The "paradoxical" results of kinematics in the STR are very well known and quoted in popular books. Here we mean relativity of scale lengths ("contraction" of lengths), relativity of simultaneity and distances between events, and relativity of time intervals between events. All these conclusions deviating from the "generally accepted" classical results were discussed in detail in §§ 3.1-3.3. In the next sections we shall consider the paradoxes which are further removed from the STR fundamentals.

§ 8.1. Faster-than-light velocities. As we saw in § 3.4 the principle of causality requires the signal transmission velocity, that

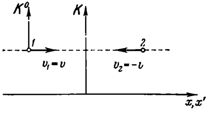


Fig. 8.1. Two particles moving toward each other can "approach" at the velocity exceeding c.

is the velocity of transmitting an energy and a momentum, to be finite. The motion of any particle whose rest mass differs from zero is, in fact, a signal since such a particle (when moving) carries along energy and momentum. Hence, it is clear that the motion velocity of such particles cannot exceed the velocity of light *in vacuo*. Kinematics of the theory of relativity shows that if in a given IFR the velocity of a particle $v < c$, then in any other IFR K' its velocity $v' < c$ (see § 3.5). Let us consider another useful example in this connection.

Let two particles move toward each other in the frame K at equal velocities (Fig. 8.1). The only condition that the STR imposes on the velocities of these particles is that they should be less than c. What is the relative velocity of these particles in K? Let the velocity of particle *1* (denoted by v_1) be equal to v; then the velocity v_2 of particle *2* is equal to $-v$. The relative velocity of the particles $v_{rel} = v_1 - v_2 = v - (-v) = 2v$. Whence it follows that if $v > c/2$, then $v_{rel} > c$. Can this velocity mean that a signal is travelling faster than light?

In the considered case v_{rel} is the velocity at which the distance between the particles decreases. This distance actually decreases

at the velocity exceeding that of light. No "signal", however, can be transmitted at such a velocity.

To determine the possible velocity of a signal transmission, we shall do as follows. An observer located at particle *1* wants to transmit a signal (information) by means of particle *2*. At the moment when the particles come alongside each other he delivers an "information package" to particle *2*. But the velocity at which the information leaves particle *1* is not at all the velocity at which the distance between the particles varies in K, but it is the velocity of particle *2* relative to particle *1*. In other words, the velocity at which information (a signal) is transmitted is the path covered by an information carrier per unit of time. The distance *per se* cannot be an information carrier.

Therefore, in order to calculate the velocity of a signal transmission, one has to calculate the velocity of particle *1* relative to particle *2* (or vice versa). For this purpose, let us fix a frame K^0 to particle *1*. Assuming $V = v$, we obtain $v_1' = 0$ from the general formula

$$v' = \frac{v - V}{1 - vV/c^2}.$$

This is an obvious result, of course: K' coincides with particle *1*. As to v_2', it is calculated as follows:

$$v_2' = \frac{-v - v}{1 + v^2/c^2} = -\frac{2v}{1 + v^2/c^2} = -\frac{2v/c}{1 + v^2/c^2} \cdot c.$$

This is the relative velocity of the considered particles. The last link of the equality is written down to demonstrate that $v_2' < c$. The demonstration is presented below for the general case.

Let in the frame K the velocities of particles flying toward each other be equal to $v_1 = \alpha_1 c$, $v_2 = \alpha_2 c$; the theory of relativity requires only that conditions $\alpha_1 < 1$ and $\alpha_2 < 1$ be satisfied. The case when $\alpha_1 < 0.5$ and $\alpha_2 < 0.5$ is not of much interest since even in K there is no velocity exceeding c. Suppose $\alpha_1 + \alpha_2 > 1$.

Let us introduce the frame K' where $v_1' = 0$; we have already seen that in this case $V = v_1$. Then

$$v_2' = \frac{v_2 - V}{1 - v_2 V/c^2} = \frac{-\alpha_2 c - \alpha_1 c}{1 + \alpha_1\alpha_2} = -\frac{\alpha_1 + \alpha_2}{1 + \alpha_1\alpha_2}\, c.$$

Let us prove that $\alpha_1 + \alpha_2 < 1 + \alpha_1\alpha_2$. We shall rearrange this inequality by transferring all of its terms to the right-hand side: $\alpha_1\alpha_2 - \alpha_1 - \alpha_2 - 1 > 0$. Grouping the terms, we get

$$(\alpha_1 - 1)(\alpha_2 - 1) > 0,$$

which is correct since the condition imposed on α_1 and α_2 is satisfied. The case considered earlier corresponds to $\alpha_1 = \alpha_2$.

Thus, in a given IFR a conventional particle ($m \neq 0$) cannot be accelerated to the velocity c. Nor this velocity can be reached as a result of a transition from one IFR to another. But is it still possible to find in nature velocities exceeding the velocity of light?

The first example suggests itself immediately. Let us take a solid, absolutely rigid rod (body) and push it. Its both ends will simultaneously start moving which means that a signal will be transmitted instantly. However, here the initial assumption is erroneous. There are no absolutely rigid bodies in nature. All bodies are similar to coilsprings of different rigidity. The transmission of a momentum (impact or push) from one end of a body to the other is accomplished in the form of motion of an elastic wave. And the velocity of elastic waves in solids is far less than the velocity of light. Thus, the STR stresses once more that absolutely rigid bodies do not exist in nature. By the way, an instantaneous change of a momentum requires an infinite force, even in the framework of Newton's mechanics.

Whereas the STR explicitly limits the velocity of signal transmission, no restrictions are imposed on the velocities which are not associated with the signal transmission and therefore they can exceed c. Usually a paradox crops up when a certain velocity exceeding c is found and claimed to be that of signal transmission. In the final analysis, it can always be demonstrated that the considered velocity has no relation to the signal transmission. We shall discuss a few examples now.

The straight line AB moves parallel to itself at the velocity V_1 directed normally to AB, and the straight line CD also moves parallel to itself at the velocity V_2 directed normally to CD. The angle between the straight lines is equal to θ. What is the displacement velocity of the intersection point M of these straight lines (Fig. 8.2)?

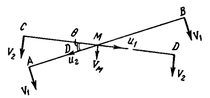


Fig. 8.2. The intersection point of two moving straight lines can move faster than light.

The relative velocity of the point M along the straight line AB due to the motion of the straight line CD is equal to $u_2 = V_2/\sin\theta$. The velocity of motion of the point M along the straight line CD due to the motion of the straight line AB is equal to $u_1 = V_1/\sin\theta$. The geometrical summation of the velocities u_1 and u_2

yields (Fig. 8.2)

$$V_M = \frac{1}{\sin\theta}\sqrt{V_1^2 + V_2^2 - 2V_1V_2\cos\theta}.$$

This formula shows that if $\theta \to 0$, the velocity $V_M \to \infty$, i.e. it can exceed c. This fact, however, by no means contradicts the theory of relativity. In the first place, the intersection point of the lines is not a material body. Secondly, this point cannot be utilized to transmit a signal (information) because at any given moment it is being formed by the new points of the two lines, i.e. the intersection point cannot be "marked".

The case of an oblique incidence of a plane light wave on a plane surface (Fig. 8.3) is of somewhat greater interest. Consider the point of intersection of the wave front with the plane $x = 0$ (the point A in Fig. 8.3). In the course of time this point moves to the right. It is easy to find the velocity of its displacement: having chosen the section BD equal to c, we obtain $AD = c/\sin\theta$. But AD is just the path covered by the point A per unit of time, i.e. the velocity of the point A. Since $\sin\theta \leqslant 1$, this velocity can be easily made greater than c. To dramatize the situation, let us imagine the plane $x = 0$ covered with a luminescent paint. Then a luminous point will run along the axis at a faster-than-light velocity.

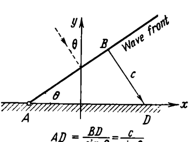


Fig. 8.3. The point of contact of an incident electromagnetic wave with a plane surface can move at a faster-than-light velocity.

Surely, a luminous point moving at the velocity $v > c$ can be realized simpler, so to speak, "manually". Arrange electric bulbs along the x axis and switch them on one after another (independently) from left to right with a given time lag. Naturally, you can get a luminous point moving at any velocity. From this second example it is seen that in this process no information can be transmitted since each source radiates independently. But is it possible to attain faster-than-light velocities by means of a relatively slow rotation of a solid body of a considerable radius? For example, a disc of radius $r = c$ rotating at the angular velocity $\omega \approx 1$ would possess the linear velocity $v \approx c$ and over at its rim. However, such a velocity cannot be reached due to relativistic properties of the motion equation. As the linear velocity of some sections of a body increases, the forces required to accelerate these sections become greater and greater, and consequently the linear velocity

of the farthest sections of the body cannot exceed c in this case either.

If it is impossible to rotate a solid body, then let us try to rotate a light beam. Let us place a searchlight at the origin of coordinates and start rotating it at the angular velocity Ω. Let us circumscribe a stationary sphere of radius c around the origin. Then the light spot will run along the surface of this sphere at the linear velocity

$$v = \Omega c.$$

This velocity can exceed the velocity of light. The example of such a beam is provided by a rotating pulsar. The light spot of the Crab Nebula pulsar runs along the Earth surface at the velocity equal approximately to 10^{22} m/s. But as in the previous cases no signal is transmitted at such a velocity. As a matter of fact, every point

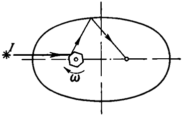


Fig. 8.4. The light spot reflected from a rotating mirror can run along the removed screen at a faster-than-light velocity.

of a screen (the Earth) receives a new portion of light energy from a searchlight (pulsar), but not from a neighbouring point of the screen. Therefore, it is impossible to transmit information from one point of the screen to another.

In fact, the same idea can also be realized as follows. A light beam from the source I falls on a mirror consisting of several facets and rotating at the angular velocity ω. Depending on the velocity ω and the distance to the screen, one can get the motion of the light spot (the source's image) at a linear velocity exceeding that of light. Let us fabricate a reflecting mirror in the shape of an ellipsoid and place a rotating mirror in one of its focal points (Fig. 8.4). Then the beam reflected from the mirror will pass through the second focal point in accordance with the well-known property of an elliptical surface. This second focal point can accommodate an analysing receiver. The light spot running along the mirror represents the image of the source regardless of the velocity of the spot.

The phase velocity of electromagnetic waves in a medium can also exceed the velocity c. It is defined by the formula $v = c/n$ where n is the refraction index. There are cases when the refrac-

tion index $n < 1$ and, consequently, $v > c$. All such cases relate to a medium and to certain frequencies of electromagnetic waves. For example, many substances have the refraction index $n < 1$ in the range of hard X-rays. The same is true for plasma. But there is no contradiction with the STR here again. The fact is that a signal transmission velocity is not defined by a phase velocity. In a dispersive medium, i.e. a medium whose refraction index depends on the frequency of light passing through it, a signal can be transmitted by means of electromagnetic waves whose frequency spectrum is sufficiently narrow (a group of waves). The velocity of a signal is the velocity of energy transmission by such a group; as a more detailed analysis (see [36]) shows, the velocity of energy transmission is defined by the group velocity. But the group velocity always turns out to be less than c, with the exception of the anomalous dispersion region where the group velocity formally exceeds c. In this region, however, the concept of a group velocity, and consequently the signal transmission velocity, loses its meaning. Thus, with the aid of wave processes, a signal is actually always transmitted at the velocity less than c.

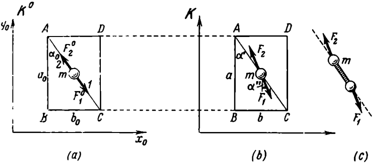


Fig. 8.5. A rectangular loop with an elastic thread stretching a sphere along the frame's diagonal. (a) The picture observed in the proper reference frame K^0; (b) the same picture observed in the frame K; (c) when the sphere is replaced by a dumb-bell, it experiences the action of a couple in terms of the reference frame K.

§ 8.2. The thread-and-lever paradox. Let a plane rectangular loop $ABCD$ be at rest in its proper reference frame K^0. An elastic thread stretches a sphere of the mass m from two sides along the diagonal AC (Fig. 8.5a). In the frame K^0 the direction of the thread is found from the triangle ABC. Designating $AB = a_0$ and $BC = b_0$, we have

$$\tan \alpha_0 = b_0/a_0.$$

Since in the frame K^0 the elastic forces are directed along the thread, we can also write that

$$\tan \alpha_0 = b_0/a_0 = F^0_{1x}/F^0_{1y}, \tag{8.1}$$

where $\boldsymbol{F}^0_1$ denotes the force directed toward the apex C. Similar relations are valid for $\boldsymbol{F}^0_2$ as well.

Now let us pass to the frame K relative to which the frame K^0 moves at the velocity V. As usual, we assume that the x_0 and x axes coincide and that the y_0, y and z_0, z axes are respectively parallel. According to the formulae (3.5) and (5.34) of length and force transformation we get

$$a = a_0, \qquad b = b_0 (1 - \mathrm{B}^2)^{1/2}, \tag{8.2}$$

$$F_{1x} = F^0_{1x}, \quad F_{1y} = F^0_{1y} (1 - \mathrm{B}^2)^{1/2}. \tag{8.3}$$

It is seen from this that Eq. (8.1) does not hold any more; in the frame K the angle defining the thread direction and the angle defining the direction of the forces are by no means equal:

$$\tan \alpha' = b/a = (b_0/a_0) \left(1 - \mathrm{B}^2\right)^{1/2} = \Gamma^{-1} b_0/a_0 = \tan \alpha_0/\Gamma, \tag{8.4}$$

$$\tan \alpha'' = F_{1x}/F_{1y} = \left(F^0_{1x}/F^0_{1y}\right)/\left(1 - \mathrm{B}^2\right)^{1/2} = \Gamma F^0_{1x}/F^0_{1y} = \Gamma \tan \alpha_0. \tag{8.5}$$

Although the sum of the forces remains equal to zero as before, these forces in the frame K are directed at a certain angle to the thread (Fig. 8.5*b*). At first glance this circumstance seems surprising. Indeed, what happens, for example, if we cut the thread at the section *2*. In the frame K^0 the acceleration must be parallel to the force direction at the initial moment, i.e. it is directed along the thread (since this is an explicitly non-relativistic case, the conventional law of Newton is quite applicable). It seems that in the frame K the acceleration should be directed at a certain angle to the thread since the direction of the thread and the direction of the force $\boldsymbol{F}_1$ do not coincide. These statements are clearly contradictory, but the paradox is resolved simply: in relativistic dynamics the acceleration does not, generally speaking, coincide with the acting force direction and although the force is directed at an angle to the direction of the thread, the acceleration is oriented along the thread. The paradox itself provides a useful illustration of the characteristics of the relativistic equation of dynamics.

Let us make sure that in both frames the acceleration of the sphere is directed along the thread. It is convenient to write the relativistic equation of motion in the form

$$m\, d\boldsymbol{v}/dt = \gamma^{-1} [\boldsymbol{F} - (\boldsymbol{v}/c^2)(\boldsymbol{F}\boldsymbol{v})];$$

here m is the mass of the sphere, $\boldsymbol{F}$ the conventional three-dimensional force acting on the sphere, $\boldsymbol{v}$ the velocity of the sphere, $\gamma = (1 - \beta^2)^{-1/2}$ where $\beta = v/c$.

In the frame K^0 at the moment $t = 0$ when the thread 2 is cut

$$m\, d\boldsymbol{v}^0/dt = \boldsymbol{F}_1^0,$$

or, in components,

$$m\, dv_x^0/dt = F_{1x}^0, \quad m\, dv_y^0/dt = F_{1y}^0.$$

Having divided the first relation by the second one termwise, we obtain the formula defining the direction of motion at the initial moment:

$$dv_x^0/dv_y^0 = F_{1x}^0/F_{1y}^0 = \tan\alpha_0.$$

In accordance with Eq. (8.1) this acceleration direction coincides with the direction of the thread, as it should be. Thus, in K^0 the forces and the acceleration are parallel, and the motion is directed along the thread at the initial moment.

Now let us pass over to the frame K. In this frame the sphere moves at the velocity coinciding with that of the frame K^0, i.e. V. Therefore $\gamma = \Gamma$ and the acceleration components can be written here as follows:

$$m\, dv_x/dt = [F_{1x} - (V/c^2)\, F_{1x} V]/\Gamma = F_{1x}/\Gamma^3, \tag{8.6}$$

$$m\, dv_y/dt = F_{1y}/\Gamma; \tag{8.7}$$

we have taken into account here that the velocity of the sphere coincides with the velocity of the frame K, i.e. is equal to V and has the components $(V, 0, 0)$; F_{1x} and F_{1y} are the force components in the frame K *. In order to find the acceleration direction in K, we shall divide Eq. (8.6) by (8.7) termwise:

$$dv_x/dv_y = (F_{1x}/F_{1y})/\Gamma^2 = \Gamma\tan\alpha_0/\Gamma^2 = \tan\alpha_0/\Gamma = \tan\alpha'. \tag{8.8}$$

In the third link of this chain of equations we utilized the formula (8.5) and in the last link the formula (8.4). But we see from Eq. (8.8) that the acceleration in K at the initial moment is also directed along the threads, and therefore no paradox can arise.

Let us imagine, however, that instead of the sphere that was implied to be a point, the threads stretch some solid object, e.g.

* It is easy to notice that Eqs (8.6) and (8.7) correspond to two exceptional cases of the relativistic equation when the force and acceleration are parallel; formerly the corresponding masses were referred to in this case as "transverse" and "radial" masses Then these rather incongruous terms have been practically discarded although they convey the tensor character of the relationship between the force and acceleration in relativistic mechanics quite satisfactorily.

a dumb-bell. Then in the frame K the spheres of the dumb-bell would be subjected to the couple (Fig. 8.5c) and the dumb-bell would be shifted relative to the diagonal of the loop.

It is obvious, however, that in the proper frame the dumb-bell axis coincides with the diagonal. Certainly, we come across a paradox here. This paradox, however, represents a version of the well-known paradox of the lever which we shall discuss now. Let a lever be at rest in the frame K^0 (Fig. 8.6). It is in the state of equilibrium in spite of the fact that two forces, F_x^0 and F_y^0, act on it, each of which is directed along the respective coordinate axis.

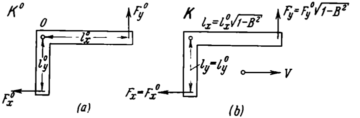


Fig. 8.6. The lever paradox. In the frame K^0 the lever is in equilibrium and the resultant force moment is equal to zero. When the same lever is considered in terms of the frame K, the force moment different from zero arises in accordance with the formulae for length and force transformations. The STR provides a very elegant explanation why the lever is at rest in terms of the frame K (see the text).

The equilibrium is ensured by the equality of the force moments in K^0:

$$F_x^0 l_y^0 = F_y^0 l_x^0; \tag{8.9}$$

these moments are oriented in opposite directions.

The same lever can be considered in terms of the frame K relative to which the lever moves as a whole at the velocity V. Having formed an expression for the moments of the forces F_x and F_y in K, we note that they are no longer equal, and, consequently, a resultant force moment acting on the lever must appear. In fact, in accordance with Eqs. (5.34) and (3.5) we have

$$F_x = F_x^0, \qquad F_y = F_y^0\sqrt{1-\mathrm{B}^2},$$
$$l_x = l_x^0\sqrt{1-\mathrm{B}^2}, \qquad l_y = l_y^0.$$

The difference of the moments of the forces F_x and F_y produces a torque in K

$$L = F_x l_y - F_y l_x = F_x^0 l_y^0 - \left(1-\mathrm{B}^2\right) F_y^0 l_x^0 = \mathrm{B}^2 F_y^0 l_x^0 = -\mathrm{B}^2 F_x l_y, \tag{8.10}$$

where Eq. (8.9) is used. Thus, the paradox consists in the following. although it is well known that the lever is motionless, in the frame K the lever is subjected to a force moment and, consequently, must be rotating. The witty solution of this paradox belongs to Laue. We have got used to a force moment inducing a rotation, or, in other words, causing the appearance of the moment of momentum in the system. In the frame K the force moment indeed defines the velocity at which the moment of momentum grows, but this growth is not associated with the rotation of the lever. Then where does the increment of the moment of momentum come from? Let us consider the work performed by the forces F_x and F_y in the frame K. In the frame K the lever moves, and the force F_x performs the work $-F_x V$ per unit of time. The force F_y performs no work since it is oriented normally to the lever velocity direction. Consequently, at the end of the lever, i.e. at the point of application of the force F_x, the work is performed, and the energy of the lever at this point increases by the quantity $-F_x V$ per unit of time. This means that the lever mass at the point of application of the force increases by $-F_x V/c^2$ per unit of time. Multiplying this quantity by the lever velocity V, we obtain the increment of the momentum $-F_x \mathrm{B}^2$. Hence, the moment of momentum grows by $-F_x l_y \mathrm{B}^2$ per unit of time. This is precisely the resultant moment cited in Eq. (8.10). So, this additional moment does not describe the rotation but defines the velocity at which the system's moment of momentum varies. This explanation has some weak points. In the STR there is no absolutely rigid bodies, and we must make allowance for the deformation of the lever. In the foregoing reasonings the lever was tacitly assumed to keep its form. In the frame K^0 we must consider the lever's arms bent by the forces F_x^0 and F_y^0.

Considering this lever, we come across still another paradoxical result. Suppose, no forces act on the lever till the moment $t = 0$ when F_1^0 and F_2^0 are "switched on" simultaneously in the frame K^0. At every moment of time the equilibrium in the frame K^0 will be maintained. In the frame K, however, the forces will not be switched on simultaneously since there will be a time interval when the force F_1 is already acting and the force F_2 has not been "switched on" yet. Consequently, a force moment arises. The following simple example shows that the forces applied at different points of the body are indeed essential. (The paradoxes appear, of course, when a solid body is considered) Let in the frame K^0 a solid body of the length l^0 be located along the x^0 axis. No forces act on this body until the moment $t = 0$ when the oppositely directed equal forces are switched on at both sides. In the frame K^0 the equilibrium is permanently maintained while in the frame K there is a time interval during which the forces are not balanced and consequently

the body should start moving. We shall leave this paradox for the reader to analyse.

§ 8.3. The tachyons. This is the name for the particles whose velocity exceeds that of light *in vacuo*. From the very beginning we should notify the reader that we speak of hypothetical particles: the experimental attempts to observe such particles have not succeeded so far. But the very idea of their existence seems paradoxical: the finite velocity of signal transmission is fundamental for the STR, the ultimate value being that of the velocity c. Surely, velocity *per se* has no limitations whatever (see § 8.1), but the signal transmission is the propagation of energy and momentum. The motion of particles to which we have got used can positively serve as a signal. Besides, the conventional particles possessing a finite rest mass, with which we have made ourselves familiar, cannot reach the velocity of light. From the relativistic equation of motion for such particles it follows that the velocity of light can be reached only after an infinitely long time (not to mention the fact that an infinitely high energy would be needed in that case). Thus, the question about the faster-than-light velocity of particles in our conventional world no longer arises.

One may, however, assume the existence of a special group of particles whose conversion into conventional particles and back is impossible. These particles, possessing faster-than-light velocities from the very beginning of their existence could have been generated in certain nuclear transformations. The assumption concerning the generation of tachyons was evoked by the picture of photon generation: at the very beginning photons possess the velocity of light, and do not emerge "dynamically", as a result of an acceleration of conventional particles.

As in § 3.5 it can be shown that if the velocity of a particle v exceeds c in one IFR, this is true in any other IFR. Consequently, the conventional particles (photons) and tachyons form separate groups of particles; the transition from one group of particles to another by means of acceleration is impossible; the transition from one IFR to another leaves a particle in the same group to which it belonged in the initial IFR.

Assuming the existence of such particles, let us consider the kinematic consequences of such an assumption.

So, let us assume that the velocity of a tachyon v determined by the conventional means exceeds c, i.e. $\beta = (v/c) > 1$. Then for the interval between two events, the positions of a tachyon at two points in space at two moments of time, we get, as usual

$$ds^2 = c^2 dt^2 - dx^2 = c^2 (1 - \beta^2) dt^2.$$

Here we consider the motion along the x, x' axis. As distinct from the conventional particles, for a tachyon $ds^2 < 0$, i.e. the in-

terval is space-like. We saw in § 3.4 that in that case the concepts "later" and "earlier" for two events are no longer absolute. Consequently, there are such reference frames in which a tachyon moves in one direction, and there are others in which it moves in the opposite direction. One can find the condition making the velocity of a tachyon reverse its direction in a certain reference frame K'. In the frame K' we obtain for a tachyon

$$\Delta t' = \Gamma\left(\Delta t - \frac{V}{c^2}\Delta x\right) = \Gamma(1 - \beta B)\,\Delta t.$$

In any IFR we assume $B < 1$. The time intervals $\Delta t'$ and Δt have opposite signs and this means that the sequence of events varies

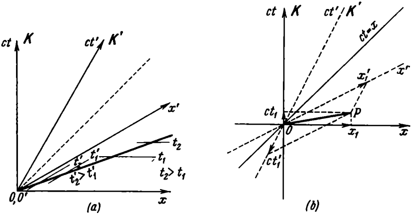


Fig. 8.7. (a) The motion of a tachyon considered in two IFRs. In the frame K a tachyon moves to the right, in K' to the left. The bold line represents the world line of the tachyon. (b) The reversal of the time sequence of events for a moving tachyon.

with time, provided $1 - \beta B < 0$. From this the required condition $v > c^2/V$ follows; it is clear that $v > c$. The differences in describing the motion of a tachyon in the frames K and K' are clearly seen in Fig. 8.7a. In the frame K the simultaneity lines are parallel to the x axis, and drawing them farther and farther from the positive ct axis, we mark the position of the tachyon more and more to the right: the tachyon moves to the right. In the frame K' the simultaneity lines are parallel to the x' axis. Drawing these lines till they intersect the ct' axis farther and farther along the positive direction of the ct' axis, we find the tachyon located more and more to the left: the tachyon moves to the left.

The same result can be presented in a more dramatic manner (Fig. 8.7b). Let in the frame K a tachycn that left the point O

arrive at the world point P. In the frame K, as it is seen in the figure, a tachyon was "emitted" at the moment $t = 0$ ("earlier") and arrived at the point P at the moment t_1, i.e. "later". The same figure features the spatial and temporal axes of the frame K' in which the simultaneity lines are parallel to the x' axis. It is seen in the figure that the tachyon in the frame K' was earlier at the point P (at the moment $-t_1'$), then moved to the point O to be absorbed at the moment $t' = 0$. In this way, we can obtain the movement of the tachyon in the opposite direction in space only

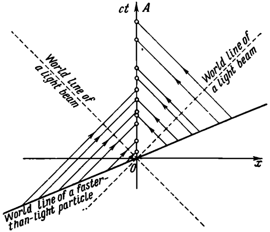


Fig. 8.8. The observation of a luminous particle moving at a faster-than-light velocity.

by a proper choice of the reference frame. As a result, in a certain reference frame we can observe the absorption of a tachyon instead of its emission.

At the same time, we shall mention a curious picture of a "luminous tachyon", that is a tachyon which radiates light. Fig. 8.8 illustrates that an observer at rest in the frame K will "see' two such tachyons diverging in the opposite directions.

Now let us go back to the reversion of the sequence of events in time and, in particular, to the "emission" and "absorption" exchange. At first glance, this situation contradicts the conventional cause-and-effect relations. Indeed, suppose it is known that the source of tachyons is located at the point O. The source is the "cause" of a tachyon generation. The motion of the tachyon toward P is the "effect" of the tachyon generation. The observation in the frame K' shows, however, that the tachyon leaves the point P and is absorbed at the point O. No matter how strange it may seem, it must be admitted that the observed sequence does not contradict

the cause-and-effect relations, provided we define clearly what we mean by them. For example, one may argue as follows.

We shall take A for the cause and B for the effect, provided that a repetition of the event A at the moments $t_1, t_2, \ldots$ chosen at will leads unfailingly to the occurrence of the event B at the moments $t_1 + T, t_2 + T, \ldots$. The essential point here is the controlled repetition of the event A and its correlation with the event B. In this sense the cause-and-effect relations do not depend on which event occurs earlier and which event occurs later. The sequence of events in time is not involved in the definition of a cause-and-effect relationship and cannot be used in order to differentiate between the cause and effect.

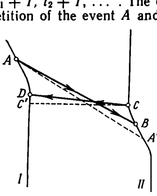


Fig. 8.9. The closed cause-and-effect cycle involving faster-than-light signals. Lines I and II are the world lines of two reference frames. The first faster-than-light signal AB is sent from the point A; AA' is the simultaneity line A faster-than-light signal CD is sent in return from the point C of frame II which arrives at the point D of frame I before the first signal was sent (the point A). The simultaneity lines and the world lines of faster-than-light signals are drawn in accordance with Fig 2.6*b*

In our example the event to be controlled in the frame K' is the absorption of a tachyon. This controlled absorption will always be preceded by the motion of the tachyon from the point P toward O. We shall have to take the absorption of the tachyon for the cause and its motion for the effect. The cited definition of the cause and effect conflicts with the conventional statement that "the absolute meaning of the notions 'earlier' and 'later' ... is a requisite condition for the concepts 'cause' and 'effect' to make sense". Of course, if the "cause" and "effect" happen at one point in a given IFR, the cause must precede the effect. But then the interval between events is time-like *a fortiori*, and in any IFRs the effect will happen "later" than the cause. The tachyons behave quite differently. All "events" involving tachyons happen at different points, when considered from our point of view. The reversal of the sequence of events is of no significance.

Hence, the reversal of the sequence of events in time does not contradict the conventional notions of the cause-and-effect relationship. However, there is one condition that had to be satisfied on all counts. It consists in the fact that it is impossible to exert influence on the past from the present. A signal sent from a given point in space cannot get at it *before* it was sent.

If tachyons served as signals, it could be possible, as it is seen from Fig. 8.9, to send a signal so that another signal caused by

the former one will get at the initial point (the cause-and-effect cycle) before the first signal was sent. Fig. 8.9 shows the world lines of two bodies, *I* and *II*, which were at rest initially, then moved uniformly and rectilinearly at equal velocities and finally came to rest again. The world points *A* and *A'* are located at the simultaneity line coinciding for both moving bodies. The world points *C* and *C'* are located at the simultaneity line coinciding for both resting bodies. The figure also illustrates the world lines of two faster-than-light signals *AB* and *CD*. Having sent the signal *AB* and then, on receiving it in another frame, the other signal *CD*, we will receive the signal *CD* at the point *D* before the signal from *A* was sent.

Thus, we have analysed the example of a closed cause-and-effect cycle indicating explicitly the possibility of exerting influence on the past. Certainly, this result pertains to any faster-than-light signals, but when applied to tachyons, it implies that tachyons themselves, unlike the conventional particles, cannot serve as signals.

If one assumes that tachyons exist and the requirements of the cause-and-effect cycle are satisfied, the resulting possibility of reversing the sequence of events in time for tachyons allows the objections concerning the "dynamic" properties of these particles to be discarded. If one assumes the basic relations of the STR to hold for tachyons, the transformation formulae for the velocity and energy of a particle (see Chapters 3 and 5)

$$\beta' = \frac{\beta - B}{1 - \beta B}, \quad \mathscr{E}' = \Gamma \mathscr{E} (1 - \beta B)$$

show that the tachyon's energy becomes negative in those reference frames where the sequence of events reverses its order and where the sign of the velocity changes to the opposite since Δt and $\Delta t'$ are of opposite sign. The negative energy of a tachyon is inadmissible since its existence would imply the possibility of obtaining unlimited energy. In fact, the joint generation of a pair of tachyons, one possessing the negative and the other the positive energy, would not require any energy expense and the positive energy tachyon could perform useful work.

We have seen, however (see Fig. 8.8), that if in the reference frame *K* a tachyon is emitted and absorbed, in the frame *K'* where the tachyon's velocity obeys the condition $v > c^2/V$ the same process can be described as an absorption of a tachyon moving in the opposite direction and possessing the positive energy. This circumstance makes it possible to avoid the difficulty associated with the emergence of negative energies.

And, in conclusion, a few remarks concerning the momentum and energy of tachyons. In a unidimensional case ($p_x = p$) the

STR yields (see Chapter 5) the following equation.

$$\mathscr{E}^2 - p^2c^2 = m^2c^4. \quad (8.11)$$

Plotting $\mathscr{E} = \mathscr{E}(p)$, we get a hyperbola; besides, as we have seen in § 5.5,

$$v = d\mathscr{E}/dp. \quad (8.12)$$

If the particle accelerates, it moves along the hyperbola (8.11) in the plane $(\mathscr{E}, p)$. The tangent slope is always less than c, irrespective of the manner in which the particle's energy increases: whether via acceleration or due to the transition to another reference frame. Since the particle's energy is positive, the lower branch of the hyperbola is disregarded. Note also that the hyperbola's asymptotes $\mathscr{E}^2 - p^2c^2 = 0$ correspond to photons. If one assumes tachyons to obey the basic formulae of relativistic mechanics (see Chapter 5), the quantities $\boldsymbol{p} = m\gamma\boldsymbol{v}$ and $\mathscr{E} = mc^2\gamma$ become imaginary since $\gamma = (1 - \beta^2)^{-1/2}$ and $\beta > 1$, so that $\gamma = i\gamma_*$ where $1/\gamma_* = \sqrt{\beta^2 - 1}$. The real values of momentum and energy can be obtained, provided we take the quantity im_* for mass. Why is an imaginary mass any better than an imaginary energy and momentum? The point is, m_* is the imaginary proper mass of a tachyon, and there is no reference frame in which a tachyon could be at rest (a reference frame consists of conventional particles and its velocity never exceeds c). Therefore the tachyon's proper mass cannot be observed and can be assumed to possess any magnitude.

But then in the plane $(\mathscr{E}, p)$ we must consider two more hyperbolas corresponding to the imaginary proper mass $z^2 - p^2c^2 = -m_*^2c^4$. Consequently, we must analyse three hyperbolas in the plane $(\mathscr{E}, p)$ (Fig. 8.10). The slope of the tangent of these hyperbolas is more than c at any point. Of course, the factor γ appears not only in the expressions for a momentum and an energy, but also in the definition of length via the proper length and the definition of time intervals via the proper time. However, we can readily reject the "proper" quantities, assuming them unobservable.

Having referred the reader elsewhere for details, let us summarize. In the past few years some attempts were made to analyse the properties of faster-than-light particles in terms of the STR. The STR says that the velocity that cannot be associated with the real physical propagation of anything can have any magnitude. Conventional particles always move at the velocity less than c, that is any "signal" propagates at a velocity less than c. Consequently, a tachyon cannot serve as a signal, i.e. its interaction with our world is quite restricted. It might be possible that the interaction of tachyons with our world is accomplished through the exchange of electromagnetic signals.

Proceeding from the motto "everything that is not forbidden has the right to exist", we must admit the existence of tachyons. So far, in theory their existence is not directly prohibited. Still it seems highly unlikely that such particles really exist. The last word is with an experiment.

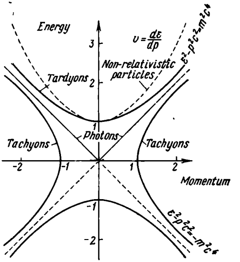


Fig. 8.10. Tachyons and conventional particles depicted in the plane ($\mathscr{E}$, p).

§ 8.4. The clock paradox. This paradox, provided there is a paradox here at all, arises due to the difference in measurements of time intervals between events in different IFRs, which was repeatedly discussed. We shall recall briefly the results that will be needed later.

Let a body be at rest in the frame K' and two events be registered at the point x' at the moments t_1' and t_2' by the clock moving together with that body and the frame K'. The interval $t_2'-t_1'$ is the proper-time interval to be denoted by $\Delta\tau$. The same two events will be registered by the observers of K at two points of the frame K by two clocks at the moments t_1 and t_2. The time interval between the same two events will turn out to be equal to $\Delta t = t_2 - t_1$. We know that

$$\Delta\tau = \sqrt{1-\beta^2}\,\Delta t, \qquad (8.13)$$

i.e. the proper-time interval between events is less than the interval between the same events registered by the clock of the frame relative to which the body moves (cf. § 3.3).

Eq. (8.13) clearly shows an asymmetry in time readings. It seems that we could argue as follows. Since all clocks in K are

synchronized, the time interval Δt measured by *different* clocks of K can be equated to the time interval registered by one clock of K. Then it will turn out that the rates of the identical clocks in two IFRs, K and K', are different. But the STR is based on the complete symmetry of inertial frames! And this symmetry does exist! We have missed an important point in our reasoning. Since the simultaneity is relative, the clocks synchronized in one frame are not synchronized in terms of another frame. The clock synchronization is relative! The quantity Δt is by no means the proper-time interval for a clock of the frame K. Let us make the requisite calculations.

Let clock *III* be at rest at the origin of the frame K' moving relative to K at the velocity V. Clock *I* synchronized in the frame K rests at the point $x_1 = a$ and clock *II* synchronized in the same frame rests at the point $x_2 = b$.

The variable coordinate of clock *III* in the frame K is equal to $x_3 = Vt$. Consequently, the coordinates of clocks *I*, *II*, *III* in the frame K will be as follows:

$$x_1 = a \quad \text{(for clock } I\text{)}, \tag{8.14}$$

$$x_2 = b \quad \text{(for clock } II\text{)}, \tag{8.15}$$

$$x_3 = Vt \quad \text{(for clock } III\text{)}. \tag{8.16}$$

From the Lorentz transformation formula (2.11) $x = \Gamma(x' + Vt')$ we can obtain the dependence of the coordinate x' in the frame K' on time t' in this frame and the coordinate x in K in the following form:

$$x' = -Vt' + \frac{x}{\Gamma}.$$

In this way we shall find x_1' and x_2'; as to x_3', it is obvious that $x_3' = 0$. Consequently,

$$x_1' = -Vt' + a/\Gamma \quad \text{(for clock } I\text{)}, \tag{8.17}$$

$$x_2' = -Vt' + b/\Gamma \quad \text{(for clock } II\text{)}, \tag{8.18}$$

$$x_3' = 0 \qquad \text{(for clock } III\text{)}. \tag{8.19}$$

As usual, we assume the readings of clocks from different frames to be comparable, when the clocks are located at one point. Then the following intercomparisons can be actually carried out. First, the reading of clock *III* can be compared to that of clock *I* when the clocks are passing each other; we shall denote the respective readings of the clocks by t_1' and t_1; secondly, the readings of clocks *III* and *II* can be compared when those clocks pass each other; we shall denote these readings by t_2' and t_2 (Fig. 8.11). When the reading of clock *III* coincides with that of clock *I*, both

clocks are located at the point $x' = 0$; therefore, in accordance with Eq. (8.17) the time moments t_1' and t_2' will be given by the following relations: $t_1' = a/V\Gamma$ and $t_2' = b/V\Gamma$. At the same time from Eq. (8.16) we obtain $t_1 = a/V$, $t_2 = b/V$, i.e. $t_1 = \Gamma t_1'$ and $t_2 = \Gamma t_2'$. The readings t_1 and t_2 are the readings of different clocks synchronized in the frame K.

Due to the synchronization in this frame clock I showed the moment t_2 when clock II showed the same moment t_2. The difference $t_2 - t_1$ is the time in the frame K during which the reading of clock III changed by $t_2' - t_1'$. In terms of the frame K the rate of clock III is defined by the following relation:

$$t_2' - t_1' = \frac{1}{\Gamma}(t_2 - t_1) =$$
$$= (t_2 - t_1)\sqrt{1 - \beta^2}, \quad (8.20)$$

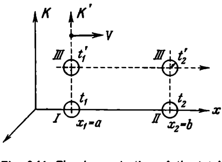


Fig. 8.11. The demonstration of the total symmetry of two inertial frames of reference with respect to time "deceleration". In any reference frame the proper-time interval between two events will turn out to be less than the time interval between the same two events registered by two clocks of any other IFR.

as it should be, since $t_2' - t_1'$ is the proper-time interval. Since $|t_2' - t_1'| \leqslant |t_2 - t_1|$, the moving clock observed from the frame K is slow. We are well aware of all this. Now we pass over to taking a decisive step: we need to compare the rates of clocks I and II as observed from the frame K'. To pass judgement on the clock rate, one should analyse the rate of one of the clocks, say clock II. However, we have only one direct reading of this clock: when clock II was against clock III, the latter showed t_2' and the former showed t_2. The other reading of clock II has to be calculated (cf. § 2.4). We shall find where clock II was located and what it showed at the moment clocks III and I were against each other. Now let us analyse the situation in terms of the frame K'. When clock III was against I, it showed the time $t_1' = a/V\Gamma$. Clock II was at the distance $x_2' - x_1' = (b - a)/\Gamma$ from clock I (see Eqs. (8.17) and (8.18)). But when clock I was against III, their common coordinate was $x_1' = x_3' = 0$. Therefore, $x_2' = (b - a)/\Gamma$ is the coordinate of clock II at the moment clocks I and III coincide. Now it is easy to find the reading of clock II at the same moment of time. We shall introduce $x_2' = (b - a)/\Gamma$ and $t_1' = a/V\Gamma$ into the formula

$$t = \Gamma\left(t' + \frac{V}{c^2}x'\right).$$

(Since the clocks are synchronized in the frame K', t_1' coincides with the reading of the clock of K' located at the point x_2' so that $t_1' = t_2'$.) As a result, we get the reading of clock II:

$$t = \Gamma t_1' + \Gamma \frac{V}{c^2} \frac{b-a}{\Gamma} = t_2 + \frac{V}{c^2}(b-a). \qquad (8.21)$$

Were clocks I and II synchronized; they would show the same time. But they are synchronized only in K and not in K'. We see that in terms of K' the clock of K is dissynchronized, and the following difference of readings appears

$$\delta = \frac{V}{c^2}(b-a),$$

which grows as the clocks move away from each other. We already obtained this result in § 2.4. Since in the frame K the distance $b - a = V(t_2 - t_1)$, the reading of clock II will be $t = t_1 + + (V/c)^2 (t_2 - t_1)$. Composing the difference of the marked time t_2 and the calculated time t, we shall obtain

$$t_2 - t = (t_2 - t_1) \frac{1}{\Gamma^2},$$

or, in accordance with Eq. (8.20),

$$t_2 - t = (t_2' - t_1') \frac{1}{\Gamma}.$$

This means that the observer in the frame K' will note that the clock moving relative to him is slow. Thus, the full equivalence of the frames is proved.

This result confirms the full equivalence of the two inertial frames considered: if in two IFRs there are two identical clocks, the *proper-time* intervals registered by these clocks are equal. Surely, it cannot be otherwise, since one of the basic principles of the STR is the principle of relativity: if the rates of identical clocks were different in two IFRs, such a physical method of distinguishing these frames would be possible.

This is only the introductory explanation that had to be made. The clock paradox is, of course, something different. Suppose we compare the readings of two clocks: one from the frame K and the other from K'. Naturally, immediately after the comparison the clocks will start differing from each other. Now the question arises: if we bring somehow one of these clocks back to the point where the other clock is located and intercompare their readings again, what should we expect? It is the answer to this question that is the clock paradox. This answer is far from being simple, and we wish the reader to arm himself with patience.

First of all, we should point out that all formulae of the STR involve the quantities treated in terms of inertial frames of ref-

erence. All time measurements carried out in the STR are made by means of clocks resting in one or another IFR. Having once compared two clocks, we are no more capable of getting them together at one point in space without taking them out of the reference frame in which they were at rest during the initial comparison. Indeed, if the motion is rectilinear, one of the clocks should be decelerated and then accelerated in the opposite direction to attain the same velocity. In that case the clock whose direction of motion was reversed will find itself at the same point as the clock against which the initial comparison was made. All this can be seen very well in the Minkowski diagram where the world lines of two clocks, *I* and *II*, are depicted (Fig. 8.12).

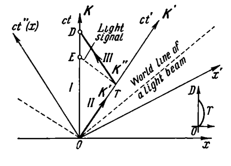


Fig. 8.12. The world lines of two clocks *I* and *II*. The world line *OD* corresponds to clock *I* resting in *K*. Clock *II* first moves uniformly from clock *I* (the line *OT*), then having altered the velocity to the equal and oppositely directed velocity at the point *T*, approaches clock *I* again. At the point *D* they get together and their readings can be compared again (the first intercomparison was made at the point *O*). This comparison of clocks amounts to what is called the clock paradox. The inset illustrates the world line of one clock getting back to the point *D*

The "clock paradox" is very convenient to analyse by means of *K* calculus (§ 3.7). We shall make use of the space-time chart of Fig. 8.12. It illustrates the world lines of three clocks: clock *I* located at the origin of *K* (the line *OD*), clock *II* resting at the origin of K' (the line *OT*) and, finally, clock *III* resting in K'' (the line *TD*). Let us find the measured time intervals directly. At the moment $t = t' = 0$, when the origins *O* and O' coincide, the initial exchange of light signals occurs which takes no time since clocks from *K* and K' are positioned at one point. Clocks *II* and *III* get together at the world point *T*; at this moment a light signal is sent from the point *T* to clock *I*. Let clock *II* register the proper-time interval $\Delta\tau_2$ between its encounters with clocks *I* and *III*. Then, as we know, clock *I* must register the time interval $k\,\Delta\tau_2$ between the encounter of clocks *I* and *II* and the arrival of the light signal from *T*. But the signal from *T* was sent at the moment when clocks *II* and *III* were against each other, and therefore if clock *III* registers the proper-time interval $\Delta\tau_3$ between the encounter of clocks *II* and *III* till its arrival at the point *D*, it is possible to find the time interval between the reception of the light signal by clock *I* at the point *E* and the encounter of clocks *I* and

III at the point D. We saw in § 3.7 that if the sign of the relative velocity of the two reference frames changes to the opposite, the coefficient k changes to $1/k$. Consequently, the time interval shown in Fig. 8.12 by the section ED is equal to $\Delta\tau_3/k$. From the symmetry of the imaginary experiment discussed here it is clear that $\Delta\tau_2 = \Delta\tau_3$. Designating the magnitude of this time interval by $\Delta\tau$, we conclude that the time interval registered by clock *I* between its encounters with clocks *II* and *III* is equal to

$$k(\Delta\tau_1) + \frac{\Delta\tau_2}{k} = (k^2 + 1)\frac{\Delta\tau}{k}. \qquad (8.22)$$

The *total* time registered by the two observers (clocks *II* and *III*) is equal to $2\Delta\tau$. This quantity is always less than the one given by Eq. (8.22) since from the inequality $(k-1)^2 > 0$ it immediately follows that

$$k^2 + 1 > 2k.$$

The obvious advantage of this approach lies in the fact that all time measurements are performed by means of clocks resting in inertial frames of reference. Thus, the time interval between events appears shorter when measured by two inertial observers, as compared to measurements made by one observer. Note that here, unlike the case when the time interval measured by one clock is compared to the time interval between the same events measured by two clocks of another IFR, one compares the time intervals measured by clocks of *three* IFRs.

So, the use of *two* clocks, *II* and *III*, has led us to the conclusion about different measurements of time intervals. Sometimes they suggest to use the same clock in the reference frames K' and K'': at the point T clock *II* is just delivered to the frame K'' to make it possible to measure the time interval in question by the same clock. This suggestion is worth dwelling upon. Although we measure the time interval between the events O and D by means of two clocks, *I* and *II*, these clocks are far from being equivalent in the considered case. When clock *II* is delivered from K' to K'', it undergoes an acceleration and gets at a non-inertial frame. Its world line transforms into a curve (see the inset in Fig. 8.12). But the inertial motion is by no means equivalent to the non-inertial one. It is quite possible that the clock which moved due to inertia registers a longer time interval as compared to the clock which participated in the non-inertial motion. There is no inconsistency here; the same conclusion is obtained from the Einstein theory of gravitation.

We have already mentioned (see § 3.3) that any acceleration, in principle, affects the clock rate Basically, the clock's rate is "correct" in inertial frames of reference. Let the world line of a

particle be curved; this means that the particle undergoes an acceleration. At any moment of the accelerated motion we can find an inertial observer moving along a tangent to the actual motion path and having the instantaneous velocity of that motion. The clock moving with an acceleration has a "correct" rate provided it coincides with the rate of the clock constructed identically, but moving together with the inertial observer in the manner indicated.

At what point of the world line does the difference between the readings of "inertial" and "non-inertial" clocks come about? From the principle of relativity it follows that the rates of the clocks constructed identically are the same in all IFRs. Whence it is clear that the difference in the readings of two clocks brought to the same point in space is caused by the clock acceleration, i.e. by the

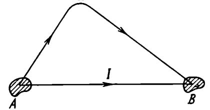


Fig. 8.13. Path *I* between towns *A* and *B* is shorter than path *II* although path *II* differs from a straight line only at a short section. The difference in length is caused not so much by a curvilinear section as by the fact that path *II* is not a straight line as a whole

curved portion of the world line. One often hears the objection that the curved portion of the world line can be made as small as needed, i.e. an acceleration can be imparted for a very short time, while the accumulated difference in time readings can be very large. We should bear in mind, however, that an acceleration imparted during a short time interval involves the immense forces, and the reversal of a relativistic velocity direction is associated with a considerable acceleration. Moreover, the difference in length between a curved world line and a straight line connecting the same points is determined not by the length of the curved portion, but by the overall curvature of the world line. This fact is very well illustrated in Fig. 8.13: although path *II* from town *A* to town *B* goes "along a straight line for almost all of the time", it is, no doubt, longer than path *I* going from *A* to *B* along a straight line connecting them. If an acceleration does not affect the clock's rate, the length of the world line of a particle determines the proper-time interval.

Until now we discussed time intervals registered by one or two clocks. Going back to the initial problem, one may ask what

clocks *I* and *III* show at the moment of their encounter at the point *D*. We remember that the sets of clocks in *K*, *K′* and *K″* are so synchronized that at the moment when the origins *O*, *O′* and *O″* of the frames coincide the three clocks from the three frames show the reading $t = t' = t'' = 0$. Now pay attention to the diagram in Fig. 8.14. Here the world lines of clocks *I*, *II* and *III* are supplemented by the simultaneity lines of the frames *K′* and *K″*. The transition from the frame *K′* to the frame *K″*, that is to another set of synchronized clocks, results in a jump of the simultaneity line from *AT* to *TB* (see the diagram). This transition explains a substantial difference in readings of clocks *I* and *III*. The substitution of two living organisms for two identical clocks leads us to the so-called twins paradox. The transition to living organisms, however, evolves a series of complications so that we refer the reader elsewhere [31].

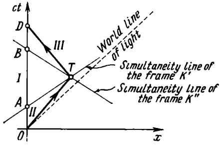


Fig. 8.14. The transition from the frame *K′* to the frame *K″* implies a change in the simultaneity line. From the line *AT* we pass over to the line *TB*.

§ 8.5. The "equivalence" of mass and energy. The zero rest mass. In this section we shall go back to the problems that have already been discussed; the main reason for this repetition is not the fact that new paradoxes will be disclosed, but that the opportunity arises to analyse jointly some results that were earlier presented separately. A few useful examples will also be given.

We know from § 5.6 that any physical system possessing the energy $\mathscr{E}_0$ in a proper reference frame ($P^0 = 0$) has a rest mass $M_0 = \mathscr{E}_0/c^2$. In this relation $\mathscr{E}_0$ denotes all the energy contained in the system. We shall illustrate this by two examples.

1. Consider a closed system consisting of *n* non-interacting material points involved in elastic collisions (in classical physics this model corresponds to ideal gas). Denote the rest masses of the points by $m_0^{(1)}, m_0^{(2)}, \ldots, m_0^{(n)}$ and the four-dimensional velocities in the proper frame K^0 by $\vec{v}_0^{(1)}, \vec{v}_0^{(2)}, \ldots, \vec{v}_0^{(n)}$. Let us pass over now to another inertial frame of reference *K* whose relative velocity direction coincides with the *x* axis. In *K* the three-dimensional velocity components of the points will be found by means of the following formulae:

$$v_x^{(k)} = \frac{v_{x,0}^{(k)} + V}{1 + \frac{v_{x,0}^{(k)}V}{c^2}}, \quad v_y^{(k)} = \frac{v_{y,0}^{(k)}\sqrt{1-\mathrm{B}^2}}{1 + \frac{v_{x,0}^{(k)}V}{c^2}}, \quad v_z^{(k)} = \frac{v_{z,0}^{(k)}\sqrt{1-\mathrm{B}^2}}{1 + \frac{v_{x,0}^{(k)}V}{c^2}}. \tag{8.23}$$

Let us also make use of Eq. (3.17):

$$\sqrt{1-\left(\frac{v^{(k)}}{c}\right)^2}=\frac{\sqrt{1-\left(\frac{v_0^{(k)}}{c}\right)^2}\sqrt{1-B^2}}{1+\frac{v_{x,0}^{(k)}V}{c^2}}, \tag{8.24}$$

which we shall rewrite using the designations adopted in this book:

$$\gamma^{(k)}=\Gamma\gamma_0^{(k)}\left(1+\frac{v_{x,0}^{(k)}V}{c^2}\right).$$

The resulting momentum of the system is defined as the sum of the momenta of individual particles:

$$\boldsymbol{P}=\sum m_0^{(k)}\gamma^{(k)}\boldsymbol{v}^{(k)} \quad \left(\boldsymbol{P}_0=\sum m_0^{(k)}\gamma_0^{(k)}\boldsymbol{v}_0^{(k)}=0\right).$$

Therefore,

$$P_x=\sum m_0^{(k)}\gamma^{(k)}v_x^{(k)}=\sum m_0^{(k)}\Gamma\gamma_0^{(k)}\left(v_{x,0}^{(k)}+V\right)=$$
$$=\Gamma\sum m_0^{(k)}\gamma_0^{(k)}v_{x,0}^{(k)}+\Gamma\frac{V}{c^2}\sum m_0^{(k)}c^2\gamma_0^{(k)}=$$
$$=\Gamma P_{x,0}+\Gamma\frac{V}{c^2}\mathscr{E}_0=\Gamma\frac{\mathscr{E}_0}{c^2}V.$$

It can be easily found that

$$P_y=P_y^{(0)}=0 \quad \text{and} \quad P_z=P_z^{(0)}=0.$$

Accordingly, the energy of the system is

$$\mathscr{E}=\sum m^{(k)}c^2\gamma_0^{(k)}=\sum m_0^{(k)}c^2\Gamma\gamma_0^{(k)}\left(1+\frac{v_{x,0}V}{c^2}\right)=$$
$$=\Gamma\sum m_0^{(k)}c^2\gamma_0^{(k)}=\Gamma\mathscr{E}_0,$$

since

$$\sum m_0^{(k)}c^2\Gamma\gamma_0^{(k)}\frac{v_{x,0}V}{c^2}=\Gamma V\sum m_0^{(k)}\gamma_0^{(k)}v_{x,0}=0.$$

Consequently, in the case of a closed system

$$\boldsymbol{P}=\Gamma\frac{\mathscr{E}_0}{c^2}\boldsymbol{V}, \quad M_0=\frac{\mathscr{E}_0}{c^2}, \tag{8.25}$$

where $\boldsymbol{V}$ is the motion velocity of the centre of inertia. This means that the rest mass of the system M_0 is equal to $\mathscr{E}/c^2$. In terms of the kinetic theory of matter the rest energy $\mathscr{E}_0$ of the system must include the thermal energy as well. However, we have already found that the rest mass of the system comprises not only the rest masses of individual particles, but also their total kinetic energy, that is the thermal energy when treated in macroscopic terms.

2. Consider an inelastic collision of two bodies. A system of two bodies can be regarded closed and therefore the conservation law for a 4-momentum can be applied to this process. Denote the rest mass of a body formed after a collision by M_0 and the rest masses of colliding bodies by $m_0^{(1)}$ and $m_0^{(2)}$. The energy-momentum conservation law will be written in the four-dimensional form as follows:

$$m_0^{(1)}u_i^{(1)} + m_0^{(2)}u_i^{(2)} = M_0 u_i, \tag{8.26}$$

where u_i is the velocity of the single body formed after the collision. The first three equations of (8.26) for $i = 1, 2, 3$ permit the three velocity components of the single body to be found. As to the fourth equation ($i = 4$), it is written as

$$m_0^{(1)}\gamma^{(1)} + m_0^{(2)}\gamma^{(2)} = M_0\gamma.$$

In the reference frame where the newly formed body is at rest

$$M_0 = \frac{m_0^{(1)}}{\sqrt{1-\left(\frac{v^{(1)}}{c}\right)^2}} + \frac{m_0^{(2)}}{\sqrt{1-\left(\frac{v^{(2)}}{c}\right)^2}}.$$

This equality can also be written as follows:

$$M_0 = \frac{1}{c^2}\left[m_0^{(1)}c^2(\gamma^{(1)} - 1) + m_0^{(2)}c^2(\gamma^{(2)} - 1)\right] + m_0^{(1)} + m_0^{(2)}. \tag{8.27}$$

It is seen from Eq. (8.27) that the rest mass M_0 of the newly formed system contains the sum of the rest masses of the initial particles $m_0^{(1)} + m_0^{(2)}$ and a certain additional mass associated with the fact that the relativistic kinetic energy of the two particles (the expression in brackets) has been transformed into some other kinds of energy (e.g. heat). Thus, in relativistic mechanics the energy conservation law includes all kinds of energy (and not only those usually taken into account in mechanics).

Finally, it should be stressed once more that the relations obtained indicate the proportionality of rest mass and rest energy; it is far more important to remember that this fact is valld only in a proper reference frame. Generally speaking, the rest energy and rest mass possess different properties under the Lorentz transformation when treated in four-dimensional terms (§ 5.7). Thus, speaking of a conversion of "mass" into energy is meaningless although sometimes one hears such a statement.

Now let us go back to a zero rest mass. Of course, from the classical standpoint a zero rest mass seems rather strange. We have seen (§ 7.6) that a zero rest mass should be attributed to the particles moving at the velocity c. In accordance with the con-

temporary ideas the particles of this kind are light quanta (photons) and neutrinos. As we know, the velocity c holds a privileged position in the STR since in all experimentally feasible IFR this velocity retains its value. We could finish here, but we would like to make some more remarks.

Obviously, there is no contradiction in regarding the matter (in its philosophical meaning) possessing a finite rest mass to be equivalent to the matter possessing a zero rest mass. We shall see that the latter case is realized comparatively rarely in nature, but basically it is feasible. These two forms of matter just mentioned can pass into one another. Now we shall dwell on one example of such a conversion. This is the formation of electron-positron pairs by gamma quanta (high energy photons) and the reverse reaction of collision between an electron and a positron (this reaction is known under somewhat obsolete name of "annihilation" of particles). This reaction brings to an end the existence of particles possessing a finite rest mass (an electron and a positron), leading to the appearance of two photons. What is essential, this reaction satisfies the momentum and energy conservation laws. Just as photons possessing a zero rest mass, so an electron and a positron possessing a finite rest mass are characterized by definite momenta and energies. The corresponding quantities resulting from this reaction remain the same; a photon as an objective reality is defined by its momentum and energy. The photon's rest mass which is equal to zero characterizes a photon none the less than a finite rest mass of an electron and a positron.

If the collision of an electron and a positron is considered in the frame fixed to the centre of inertia, that is the frame in which the particles move toward each other at equal, but oppositely directed, velocities v_1 and v_2, the energy conservation law takes the form

$$\frac{m_0c^2}{\sqrt{1-\left(\frac{v_1}{c}\right)^2}}+\frac{m_0c^2}{\sqrt{1-\left(\frac{v_2}{c}\right)^2}}=2h\nu. \tag{8.28}$$

This equation shows that the total energy of an electron and a positron is equal to the energy of two photons formed. If we take into account that in the frame of the centre of inertia $v_1 = v_2$, the observed frequency of these photons will be equal to

$$\nu=\frac{m_0c^2}{h\sqrt{1-\beta^2}}. \tag{8.29}$$

From the momentum conservation law it follows that the energies of the photons formed are equal: the momenta of the photons must be equal in magnitude (and oppositely directed), and the photon's energy is proportional to its momentum.

When an electron and a positron move at non-relativistic velocities, the frequency of photons resulting from the annihilation of such an electron and a positron is equal to $\nu = mc^2/\hbar$ which is in a good agreement with experimental data.

The example presented here is far from being unique. We may also mention the decay of a neutral (π^0) meson (possessing the rest mass equal to about 200 rest masses of an electron): $\pi^0 \to 2\gamma$.

Let us consider now n photons of the same frequency moving in various directions. The energy of this system of photons is equal to the sum of the energies of individual photons: $\mathscr{E} = \sum \varepsilon_i = nh\nu$; the momentum of the system of photons $\boldsymbol{P}$ is equal to the sum of the momenta of the photons:

$$\boldsymbol{P} = \sum \boldsymbol{p}_i = \frac{h\nu}{c}(\boldsymbol{s}_1 + \boldsymbol{s}_2 + \ldots + \boldsymbol{s}_n),$$

where $\boldsymbol{s}_i$ is a unit vector oriented along the propagation direction of the ith photon. In accordance with the definition, the rest mass M of this set of photons can be found from the expression

$$M^2c^2 = \frac{\mathscr{E}^2}{c^2} - \boldsymbol{P}^2 = \left(\frac{nh\nu}{c}\right)^2 - \left(\frac{h\nu}{c}\right)^2 (\boldsymbol{s}_1 + \boldsymbol{s}_2 + \ldots + \boldsymbol{s}_n)^2. \quad (8.30)$$

The right-hand side of Eq. (8.30) turns into zero only when all the photons propagate in one direction. This result was obtained in § 7.3· a limited train of plane waves has a zero rest mass. However, two photons whose propagation directions form a certain angle θ possess a finite rest mass. Indeed, from the general formula (8.30) we get

$$M^2c^2 = \left(\frac{2h\nu}{c}\right)^2 - \left(\frac{h\nu}{c}\right)^2 (2 + 2\cos\theta) = \left(\frac{2h\nu}{c}\right)^2 \left(1 - \cos^2\frac{\theta}{2}\right). \quad (8.31)$$

Thus, a cloud of electromagnetic radiation consisting of photons the rest mass of each of which is equal to zero possesses a positive rest mass, and, accordingly, induces a gravitational field and experiences a force of a gravitational field.

Proceeding from the fact that even two photons possess a finite rest mass, we could try to avoid discussing the zero mass of a photon. But an individual photon can be observed in principle *, and, therefore, a zero rest mass needs to be interpreted.

In order to clear up the cause that has led to the appearance of a "zero rest mass" it is expedient to use the four-dimensional concepts. Let us consider the 4-momentum of a particle of a finite rest mass m:

$$\vec{P}\left(m\gamma\boldsymbol{v}, \frac{i}{c}\mathscr{E}\right), \quad \mathscr{E} = mc^2\gamma.$$

* O. Frisch, UFN 90, 379 (1966).

The rest mass is the norm of the 4-vector $\vec{P}$:

$$\vec{P}^2 = \frac{\mathscr{E}^2}{c^2} - p^2 = m^2c^2, \qquad (8.32)$$

which is an invariant. In the 4-vector of energy-momentum the energy is represented by a time component whereas the spatial components are the components of the three-dimensional momentum. It is relevant to recall that the basic properties of the 4-vector $\vec{P}$ coincide with those of the 4-vector $\vec{V}$ since $\vec{P} = m\vec{V}$. On the other hand,

$$\vec{V} = \frac{d\vec{R}}{d\tau}, \quad d\vec{R}^2 = ds^2,$$

$$\vec{V}^2 = \left(\frac{ds}{d\tau}\right)^2. \qquad (8.33)$$

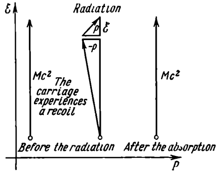


Fig. 8.15. A photon transfers mass although its mass is equal to zero. Before the radiation of a photon the energy of a carriage is equal to $\mathscr{E}_0$. In a closed system the 4-momentum is retained, so that the total 3-momentum of the carriage and a photon is equal to zero as before and the total energy of the carriage and a photon is equal to $\mathscr{E}_0$. The mass of the system remained constant although the mass of the carriage decreased while the photon's mass is equal to zero (mass is not additive!). When the photon is absorbed at the other end of the carriage, the energy of the carriage becomes equal to $\mathscr{E}_0$ again, but the energy $h\nu$ has already been transferred from one end of the carriage to the other and the mass distribution over the carriage differs from the initial one.

Therefore, in the case of the world lines of zero length ($ds = 0$) the rest mass of the corresponding particles turns into zero since $\vec{P}^2 = 0$. Thus, photons move along the lines of zero length.

There is still another question which seems paradoxical at first glance. How does a photon, whose rest mass is equal to zero, transfer a finite rest mass from one point to another? An absorption of a photon proves that this is really the case. For example, donating its energy to a solid body, a photon warms that body up and thus increases its rest mass.

Let us analyse a simple example. At one end of a carriage capable of a frictionless motion a photon is emitted; then the photon is absorbed at the other end of the carriage. Prior to the emission of the photon the energy of the stationary carriage is $\mathscr{E}_0 = Mc^2$ (Fig. 8.15). Since the system is closed, its 4-momentum remains constant and the sum of the 3-momentum of the carriage and the photon is equal to zero as before. The cumulative energy of the

carriage and the photon is $\mathscr{E}_0$. The mass of the system remains constant although the mass of the carriage decreased and the photon's mass is equal to zero. There is no reason to be frustrated about: mass is not an additive quantity! When the photon is absorbed at the other end of the carriage, the energy of the carriage is again $\mathscr{E}_0$, but by that time the energy $h\nu$ had been transferred from one end of the carriage to the other and the mass distribution over the carriage had become different from the initial one.

Finally, we want to point out that the conclusions of the STR make us define more accurately the concept of a "closed" system. In mechanics a system is called closed if the constituent bodies do not interact with "external" bodies. An interacton is described by means of forces. In chemistry they prefer to call a system closed if it does not exchange any matter with the environment (then, according to non-relativistic ideas, the mass remains constant). Passing over to thermal processes, we expect a closed system to be heat-insulated. However, the STR declares any energy transfer to be associated with a momentum transfer (this covers a heat transfer as well): an energy transfer leads to a change of a system's mass. These definitions can be combined into one, regarding a system in which an energy and momentum (a 4-vector of energy-momentum) are retained as a closed one. An energy and a momentum remain constant in a closed mechanical system. Such a system is heat-insulated. In accordance with the conventional definition $Mc^2 = (\mathscr{E}/c)^2 - \boldsymbol{P}^2$ a mass of the system remains constant. Of course, the mass conservation law, when applied to a closed system, does not imply the additivity of masses in the system. This fact should be allowed for especially in the case of generation of new particles.

SUPPLEMENT

1. Who developed the special theory of relativity, and how? * (*V. L. Ginzburg*). The theory of relativity is one of the greatest scientific discoveries of all times; moreover, it was made in our century. The latter fact is especially significant in that the theory is not so much a part of the history of science (or, if you wish, not only a part of the history of science), but a physical theory with direct and very extensive applications. That is the main reason for the heightened interest in the story of how the theory of relativity evolved. It required a reappraisal of the fundamental concepts of space and time, and thereby of the very foundations of classical (pre-relativistic) physics. Old concepts die hard and new positions are not easily won: the controversies and debates went on for decades. Moreover, they involved representatives of other sciences in addition to physicists. The theory of relativity has been and remains the focus of intense scrutiny. This, of course, is also true of its history — as an account of the evolution of ideas as well as an issue of priority.

Thus it is that even today, seventy years after the enunciation of the special theory of relativity, people are still asking, Who in fact developed it, and how?

Special relativity is most frequently associated with the name of Albert Einstein, with H. A. Lorentz, Henri Poincaré, and a few others mentioned as his predecessors. But there are other opinions which, for example, name Lorentz, Poincaré and Einstein as the joint authors of STR. What view is more justified, and what, in fact, is the argument all about? The answer to this question, as well as to the one in the heading of this supplement, should be of interest to the reader of this book. Below follow some remarks on this score.

Three works are recognized as crucial to the special theory of relativity. The author of one (1904) was the Dutch Professor Hendrik Lorentz (1853-1928), one of the leading lights in theoret-

* This is a revised version of V. L. Ginzburg's article that appeared in the final form in the 1974 *Einstein Collection* (Moscow, 1976 in Russian).

ical physics, winner of the 1902 Nobel Prize in physics. The author of the second work (1906, a brief preview of which had been published in 1905) was the celebrated French mathematician Henri Poincaré (1854-1912), also famous for his research in physics and the methodology of science. Finally, the third work (1905) was written by a virtually unknown clerk of the Swiss Federal Patent Office, Albert Einstein (1879-1955).

It is common knowledge that new works of popular and favourite writers and poets immediately attract universal attention, whereas novices have to battle against stiff odds. In science this natural tendency is, if anything, more pronounced. How come that, in the case of STR, it was the other way round, and it was Einstein's work that gained acclaim, nay, renown? A clear answer to this question was given by Wolfgang Pauli in his well-known article "Theory of Relativity", first published in 1921 in the then prestigious *Mathematical Encyclopedia*. Pauli's article was subsequently reprinted and translated into other languages (the Russian translation appeared in 1947). Pauli concludes his account of the history of the special theory of relativity with the words: "It was Einstein, finally, who in a way completed the basic formulation of this new discipline. His paper of 1905 was submitted at almost the same time as Poincaré's article and had been written without previous knowledge of Lorentz's paper of 1904. It includes not only all the essential results contained in the other two papers, but shows an entirely novel, and much more profound, understanding of the whole problem" [8]. Another eminent physicist, Max Born, recalls his impression after reading Einstein's paper: "Although I was quite familiar with the relativistic idea and the Lorentz transformations, Einstein's reasoning was a revelation to me."

It is in this entirely new and profound elucidation of the problem, making it a relevation, that the success of Einstein's work is rooted, which is what made it fundamental to the enunciation of the special theory of relativity.

A perusal of the history of science primarily focuses on two questions. The first is, How? How did ideas appear and evolve, how was a discovery prepared and made? The second question is, Who? Who made the discovery, voiced the idea, turned it into "flesh and blood", elaborated it and drove it home to the scientific community? The question How? would appear to be the basic, primary one: it is connected with the very content of science and the methods of scientific research. The question Who? may seem secondary; indeed, it has no bearing on the essence of the matter, if we take, say, physics and not the psychology of scientific creativity, the sociology of the academic milieu, or the life of this or that person. Actually, it is difficult, if not impossible, to draw the line between How? and Who? Science is advanced by people, and

if the end product — the totality of certain assertions, equations, relationships, etc. — is depersonalized or, more precisely, almost depersonalized, the initial process of the discovery of development of the equations and relationships reflects the characteristic, most typical traits of the discoverer. Thus, as far as the history of science is concerned, the questions How? and Who? must be answered simultaneously.

We shall preface further remarks on this score with a few words about the special theory of relativity (this book, of course, covers the topic in much greater detail, but it is useful to sum up the situation here).

One of the fundamental physical concepts is that of inertial frames of reference. A frame of reference used to define the coordinates and time of events is inertial if the law of inertia holds in it, namely that an isolated body (not subject to any forces) moves uniformly in a straight line. To be sure, this definition is not immune from objections and must be clarified, insofar as it remains unclear what body can be regarded as isolated? Broadly speaking, a body can be considered isolated if all other bodies are sufficiently far away. An example of a "good" inertial system is a coordinate system with the origin at the centre of the Sun and the axes directed toward the remote stars. The inertia law holds with somewhat less, but still sufficiently great, precision on the Earth (neglecting gravity). A reference system rotating relative to an inertial system is not inertial, the difference between the former and the latter being the greater the higher the angular velocity.

If a given system is inertial, any other system moving uniformly in a straight line relative to it is also inertial. The generalization of this conclusion over all mechanical phenomena — the assertion that all mechanical phenomena occur absolutely identically in all inertial systems — is just what the classical, or Galilean, principle of relativity is all about. More precisely, the definition and application of the principle incorporates the quite definite prerelativistic assumption concerning the connection between the coordinates and time of events in different inertial systems. Thus, if one such system K' (coordinates x', y', z', and t') is moving relative to a given inertial system K (coordinates x, y, z, and time t) with a velocity V along the positive axes x, x' (the direction of which we assume to coincide), then, as assumed before special relativity,

$$x' = x - Vt, \quad y' = y, \quad z' = z, \quad t' = t$$

(the Galilean transformations).

The absolute nature of time — its independence of the motion of the reference system (whence the equality $t' = t$) — was, of course, assumed to hold in all reference systems in general.

In uniform motion a body's acceleration is, obviously, zero. Hence, in the Galilean transformations, i.e. in any inertial system, the acceleration is the same. Therefore, in these transformations, the law of dynamics, Newton's second law (mass times acceleration equals force), remains unchanged as long as the mass, force, and acceleration remain the same in the K and K' systems. The latter is assumed (and proved experimentally), and we come to the conclusion that the classical principle of relativity holds in Newtonian mechanics. Generally speaking, the invariance of the equations expressing the fundamental physical laws in the Galilean transformations is proof of the validity of the classical principle of relativity.

Up till the end of the 19th century it was held that physics could be constructed completely on the basis of the Newtonian equations of motion. Thereby the classical principle of relativity was held to be invariably valid. However, the development of electrodynamics cast doubt on the classical principle of relativity. The equations of electrodynamics (Maxwell's equations) do not retain their form in the Galilean transformations, whence their application leads to the conclusion that in electrodynamics the relativity principle breaks down and, in particular, light and all other electromagnetic waves propagate differently in different inertial systems, even in vacuum. If the "luminiferous medium" — the ether — introduced then is motionless in one inertial system (K), the velocity of light in it is $c = 3 \times 10^8$ m/s irrespective of the direction. In other inertial systems K' moving with velocity V relative to the ether (along the x and x' axes), the velocity of light is, as is obvious from the Galilean transformations, $c' = c - V$ along the x and x' axes and $c' = c + V$ in the opposite direction, etc.

But experiments refuted that apparently obvious conclusion; all experiments, starting with Michelson's famous experiment performed in 1881 and repeated many times since, confirm the validity of the relativity principle in electrodynamics as well as in physics as a whole. But how, in accordance with the relativity principle, can the velocity of light be the same in different reference systems, when the Galilean transformations lead to the opposite conclusion?

It took almost a quarter of a century of agonizing quest to arrive at the solution constituting the core and basis of the STR: the Galilean transformations had to be wrong. More precisely, as is usual in such cases, they were not actually wrong, but approximate. The precise equations linking coordinates and time in the frames K' and K have the form

$$x' = \frac{x - Vt}{\sqrt{1 - \frac{V^2}{c^2}}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - \frac{V}{c^2}x}{\sqrt{1 - \frac{V^2}{c^2}}}$$

(the Lorentz transformations). If the relative velocity V of inertial systems is small compared to the speed of light c, the Lorentz transformations become the Galilean transformations; hence the degree of accuracy given by the parameter V^2/c^2. For a satellite in orbit not far from the Earth $V \approx 8 \times 10^3$ m/s, and $V^2/c^2 \approx 10^{-9}$. The velocity of the Earth around the Sun is $V \approx 3 \times 10^4$ m/s, and $V^2/c^2 \approx 10^{-8}$. It is obvious from these examples that in the domain of the phenomena we encounter in everyday life the Galilean transformations, and the Newtonian mechanics associated with them, are valid to a high degree of accuracy. But in electrodynamics, and in studying relativistic particles, travelling at velocities, v, comparable with the speed of light, c, *in vacuo*, the Lorentz transformations are required. One of their corollaries is the equation

$$x^2 + y^2 + z^2 - c^2t^2 = x'^2 + y'^2 + z'^2 - c^2t'^2.$$

Remembering that the equation of the front of a spherical light wave has the form $x^2 + y^2 + x^2 - c^2t^2 = 0$, the above equation immediately testifies to the validity of the relativity principle in the propagation of light: in all inertial systems the speed of light is the same and equal to c.

The special theory of relativity precisely represents the theoretical constructions based on the principle of relativity and the Lorentz transformations. The principal feature of the STR is the new spatio-temporal concepts as reflected in the replacement of the Galilean transformations by the Lorentz transformations. The meaning of the latter, physically speaking, is not restricted to the simple equations linking the coordinates and time x', y', z', t' with x, y, z, t. As always in physics, it is necessary to establish the meaning of all quantities, state the basis of the methods employed to measure the coordinates and time, and clarify the properties of the rulers and clocks used for this. One of the problems is that of synchronizing the clocks in each of the frames K and K'. The coordinates and time that appear in the Lorentz transformations are so defined that events simultaneous in the frame K (time t) are not simultaneous in the frame K' (time t'). The rejection of absolute time is an especially radical conclusion (for which we are indebted to Einstein). In importance and difficulty it can be compared with the rejection of the idea that the Earth was stationary on which Copernicus based his heliocentric system.

Now we can directly attack the question: Who developed the special theory of relativity, and how?

The road to the STR lay, as is apparent from what has been said, through a fundamental difficulty that had to be overcome: the principle of relativity holds experimentally in electrodynamics as well as in mechanics, but it is incompatible with the Galilean transformations. To be sure, Lorentz and others sought to remove

the contradiction without rejecting the Galilean transformations by assuming that all bodies moving with respect to the ether contract. If a ruler whose length at rest relative to the ether is l_0 is of length $l_0\sqrt{1-(V/c)^2}$ when moving at velocity V, then we can explain why some experiments do not reveal the motion of bodies relative to the ether, and their results do not depend on the velocity of the Earth's motion with respect to the Sun. However, the contraction hypothesis is not adequate for all experiments; new facts kept coming to light which agreed with the relativity principle and required additional hypotheses to explain them. This was, of course, an intolerable situation, and Lorentz stubbornly strove to show that many electromagnetic phenomena strictly, i.e. without neglecting higher order terms, do not depend upon the motion of the system. For this Lorentz had to show that for a body in *uniform rectilinear motion* (relative to the ether) the equations of electrodynamics allow for solutions which in a certain way correspond to the solutions for an identical body *at rest*. Correspondence is achieved by going over to new variables, x', y', z', and t', with the help of the Lorentz transformations, as well as the introduction of new (primed) electromagnetic field vectors. The field equations do not change as a result of these transformations, and they have the same form for the old (unprimed) and new (primed) quantities. This property is known as invariance, in the present case invariance of the electromagnetic field equations with respect to the Lorentz transformations.

Today, with special relativity, we know that this is precisely the confirmation of the validity of the relativity principle in electrodynamics, though Lorentz did not consider the time t' to be the time in the moving reference frame; he called it local time and assumed that he was dealing simply with supplementary quantities introduced by means of a mathematical contrivance. In particular, the variable t' could not be called "time" in the same sense as the variable t. In 1915 Lorentz reiterated the idea. He said that the main reason for his failure had been that he had always held that only the variable t could be taken as the true time, while his local time t' should have been regarded as no more than supplementary mathematical quantity. In Einstein's theory, on the other hand, t' plays the same part as t. In 1927, a year before his death, Lorentz stated this even more definitely, saying that for him there was only one true time and that he regarded his time transformation as merely a heuristic working hypothesis. Thus, the theory of relativity was actually the work of Einstein alone I may add that, having reread the works of Lorentz and Poincaré (70 years after their publication), it was only with difficulty, and already knowing the result (which, of course, greatly facilitates understanding) that I could understand why the invariance of the electrodynamic

equations with respect to the Lorentz transformations, proved in these works, could at the time be regarded as proof of the validity of the relativity principle.

Besides, Lorentz and Poincaré saw the principle simply as an assertion of the impossibility of observing uniform motion of a body with respect to the ether. It requires no special effort to proceed from here to treating all inertial reference frames as completely equivalent (that is the contemporary formulation of the relativity principle) only if the Lorentz transformations are understood as corresponding to going over to a moving frame of reference.

As we have seen, it was this that Lorentz definitely did not consider. Poincaré's stand is less clear. In his paper of 1906 he simply asserts that the equations of electrodynamics "can be subjected to a remarkable transformation discovered by Lorentz, the significance of which is that it explains why no experimental demonstration of the absolute motion of the universe is possible".* In my view, this "explanation" goes no farther than Lorentz's. In general, Poincaré writes: "The results which I have obtained agree with those of Lorentz in all the principal points, and I have needed only to modify and augment them in certain details. These differences, which are of but minor importance, will be shown in later sections." ** On the other hand, some of Poincaré's remarks in earlier works, papers and reports sound almost prophetic, notably regarding the need to define the concept of simultaneity, the possibility of using light signals for this, and his comments on the relativity principle. However, he did not elaborate on this, and in his works of 1905 and 1906 he followed Lorentz. As emphasized before, they strove mainly to show, and showed, under what assumptions the uniform motion of bodies relative to the ether remained undetectable. But Einstein in his 1905 work reversed, one could say, the whole issue by showing that, having accepted the relativity principle and synchronized the clocks with the help of light (and also postulating that the velocity of light does not depend on the motion of the source), no additional hypotheses were required: the Lorentz transformations follow directly from these assumptions. The contraction of moving rods and retardation of the rhythm of moving clocks can also be postulated from them.

Thus, judging by published materials, Poincaré was apparently very close to enunciating the STR, but he failed to make the final step. We can only surmise why. Perhaps it was because he was primarily a mathematician, and it was therefore especially hard

* C. W. Kilmister, *Special Theory of Relativity*, Oxford, Pergamon Press, 1970

** Ibid.

for him to rise (or descend?) to a clear understanding of such physically important aspects of the problem as adequate clarification of the meaning of all introduced quantities and concepts. Another similar hypothesis is that Poincaré was prevented by his predilection for convention, i.e. that school that emphasized (and overrated) the role of conventional elements and definitions in physics *. That convention plays a part in the development of physical theories is indubitable. Length can be measured in metres, feet or some other unusual or way-out units. The same is true of time and other quantities, as well as of the definition of simultaneity: there is no uniquely preordained definition. But the end result, the content of a physical theory (as distinct from forms of notation, etc.) is not a matter of convention, it is determined by nature, a subject of investigation. Overestimation of the conventional element in knowledge may prevent the clarification of concepts. It could, in particular, explain why Poincaré failed to clarify the meaning of "true" time, t, and "local" time, t', which are in fact equally true but are, if you care, "local" times for the frames K and K', respectively.

* As far as I can judge, these comments coincide with the view of Louis de Broglie, expressed in an address on the occasion of the birth centenary of Poincaré· "It needed but a little, and Henri Poincaré rather than Albert Einstein would have been the first to enunciate the theory of relativity in all its generalities, thereby giving French science the honour of the discovery . . However, Poincaré failed to make the decisive step, leaving to Einstein the honour of perceiving all the corollaries of the relativity principle and, in particular, through a profound analysis of the measurements of length and time, establish the real physical nature of the connection between space and time established by the principle of relativity. Why did Poincaré fail to pursue his conclusions to the end? He doubtlessly possessed an extremely critical mind, perhaps because as a scientist he was first and foremost a pure mathematician As mentioned before, Poincaré adopted a somewhat sceptical stance with regard to physical theories, holding that in general there existed an infinite number of logically equivalent points of view and pictures of reality from which the scientist, guided solely by considerations of convenience, chose one Such nominalism probably prevented him from conceding that amidst all logically possible theories some were closer to physical reality, or at least agreed better with the physicist's intuition and were therefore more useful That is why young Albert Einstein, who was only 25 at the time and whose knowledge of mathematics was in no way comparable with the great French scientist's profound knowledge, was able, before Poincaré, to find the synthesis which at once removed all difficulties, using and justifying all the attempts of his predecessors. The *coup de grâce* was dealt by a mighty intellect guided by a profound intuition of the nature of physical reality

"Yet Einstein's brilliant success should not let us forget that the problem of relativity had been earlier and profoundly analysed by the vivid mind of Poincaré, and that Poincaré made a substantial contribution to the eventual solution of the problem Einstein would never have succeeded without Lorentz and Poincaré" (L De Broglie, "Henry Poincaré, les théories de la physique." Le livre du Centenaire de la Naissance de Henry Poincaré 1854-1954, Paris, 1955)

I feel we must respect the point of view of de Broglie, whose attitude toward the memory of Poincaré was that of profound respect and maximum good-will.

I must stress, however, that such hypothetical reasoning, in this case as applied to Poincaré, is on the whole unjustified. There can be no doubt that Poincaré took an active part in the development of special relativity, and his contribution is indubitable. It is no more legitimate to ask why he failed to do Einstein's work than it is to ask the same question concerning other physicists of the time: great works are great by virtue of the very fact that they are very difficult.

In addition to what has been said of the part played by Einstein's work, here is what he himself had to say, in a letter written two months before he died: "Recalling the history of the elaboration of the special theory of relativity, it can be stated conclusively that by 1905 its discovery had been prepared. Lorentz already knew that the transformation later named after him was of key importance in analysing Maxwell's equations, and Poincaré elaborated that idea. As for me, I knew only of Lorentz's important work of 1895, but not Lorentz's later publication or the consecutive investigations by Poincaré. In this sense my work of 1905 was independent. The new element in it was the idea that the meaning of the Lorentz transformations went beyond the framework of Maxwell's equations and involved the essence of space and time. Also new was the conclusion that the 'Lorentz invariance' was a general condition for all physical theories. That was especially important to me because I had already realized that Maxwell's theory did not describe the microstructure of radiation, and therefore did not always hold." *

So, the reader wishing to receive a simple answer may ask, after all is said and done, who developed the special theory of relativity? As in most such cases, special relativity is not a discovery or result attributable solely to one person. However, most physicists (myself included) unequivocally credit Einstein with the principal role in elaborating it, for it was his work that contained an "entirely novel, and much more profound, understanding of the whole problem" [8]; it was "the last and decisive element in the foundation laid by Lorentz, Poincaré and others, on which the edifice could be built" (M. Born, Naturwiss Rundschau, 1956). First among those "others" is Larmor, who derived the Lorentz transformations back in 1900 (Vogt employed transformations very similar in form even earlier, as far back as 1887).

There are other assessments of the part played by Einstein, Lorentz and Poincaré in elaborating the STR. And whereas extremist views in effect rejecting Einstein's contribution cannot be treated seriously, more moderate statements such as "special relativity

* Quoted according to Carl Seelig's book, *Albert Einstein* (Russian edition. Moscow, 1966), the best biography of Einstein I have ever read.

was developed by Lorentz, Poincaré and Einstein" are, in the final analysis, their authors' affair: such things cannot be decreed, and no one has invented an instrument for gauging scientific merits with pharmaceutical precision.

To avoid any misunderstanding, one more comment regarding the commonly used formula "Einstein's relativity theory" is appropriate. It is a natural and legitimate formula, all the more so as it is by no means the same as saying "Einstein's special theory of relativity". For when we speak of relativity theory in general we mean both the special and the general theory of relativity. The general theory of relativity elaborates upon and advances special relativity and is generally regarded as an unsurpassed pinnacle of theoretical physics *. Max Born, for example, stated in 1955: "I have held and continue to hold that this is the greatest discovery of the human mind involving nature, which most remarkably combines philosophical depth, intuition, physics and mathematical art. I admire it as I would a work of art." Noteworthy is Einstein's own remark in a letter to A. Sommerfeld written in 1912, when he was working on general relativity: "In comparison with this problem the initial relativity theory (i.e. special relativity. — *V. G.*) was child's play." From another letter of Einstein's we know that "the period from the origination of the idea of the special theory of relativity to the completion of the paper which set it forth was five or six weeks". It took Einstein eight or nine years (from 1906 or 1907 to 1915-1916) to elaborate the general theory of relativity, after which he continued to work on it until his death on April 18, 1955. To this should be added that the general theory of relativity is, more than any other theory in the history of science, the creation of one person, Albert Einstein. Finally, relativity theory emerged from the confines of the scientific community and reached the general public only in 1919, when the deflection of light passing close to the Sun predicted by general relativity was actually observed. Hence, relativity theory as a whole can be legitimately associated with Einstein, and him alone.

Finally, a few words about who was the first. In 1952, Max Born wrote to Einstein from Edinburgh: "The elderly mathematician Whittaker, with whom I am friendly, and who resides here as honorary professor, has prepared a new edition of his old book *A History of the Theories of Ether and Electricity*, the second volume of which has already appeared. It includes, among other things, a history of the theory of relativity, with the peculiarity

* As space does not permit me to go in more detail into the place of general relativity in the development of physics, I could refer the inquisitive reader to my article "The Heliocentric System and the General Theory of Relativity (from Copernicus to Einstein)", which appeared in the *Einstein Collection 1973*, in Russian, Moscow, 1974.

that its discovery is ascribed to Poincaré and Lorentz, whereas your work is mentioned as secondary. Although the book comes from Edinburgh, I am not really afraid that you may imagine that I am behind it. In fact, for three years I have been doing all I can to dissuade Whittaker from his intention, which he cherished for a long time and which he liked to advertise. I reread the old original papers, including some of Poincaré's incidental works, and provided Whittaker with English translations of the German works. . . . But all was in vain. He insisted that everything of substance could be found in Poincaré's work and that the physical interpretation was obvious to Lorentz. But I know how sceptical Lorentz really was, and how long it took for him to become a 'relativist'. I explained all this to Whittaker, but unsuccessfully. This angers me, because he enjoys great prestige in the English-speaking countries, and many will believe him. It is especially unpleasant to me that he makes all kinds of references to partial communications regarding quantum mechanics in such a way as to especially praise my role in it. So that many (if not you yourself) may imagine that I am in some unsavory way involved in this thing." *

Einstein's answer was: "Dear Born, don't give any thought to your friend's book. Everyone behaves as seems to him right, or, expressed in deterministic language, as he has to. If he convinces others, that is their problem. At any rate, I found satisfaction in my efforts, and don't think it is sensible business to defend my few results as 'property', like an old miser who has laboriously gathered a few coins for himself. I don't think ill of him, to say nothing, of course, of you. And I don't have to read the thing." **

The answer is very typical of Einstein, and it can clarify much for those who are unfamiliar with his life. Actually, it explains the main thing: the "secret" of his exceptional popularity in the modern world. The fact that he was the greatest of the great physicists of our, and not only our, age, is fundamental, but it is not all. Einstein was also a champion of justice, freedom and other human rights, he despised the forces of evil and offered an example of nobleness and lofty human dignity. It is simply impossible to imagine Einstein engaging in arguments about, still less bickering over, priority. The same is true of Lorentz and Poincaré. Lorentz, who contributed so much to the development of special relativity, gave the credit of its enunciation "solely to Einstein", and noted Poincaré's contribution. The latter spoke highly of Lorentz's role. Einstein stressed the contributions of Lorentz and Poincaré. It

* *Albert Einstein, Hedwiga und Max Born, Briefwechsel, 1916—1955* (letter dated 26 September, 1952).

** Ibid (letter of 12 October, 1952).

could be suggested that Poincaré did not consider Einstein's contribution to be so great and perhaps even felt that he had "done it all" himself. But actually we can only speculate about what Poincaré felt from what he did not say rather than any complaints he ever voiced.*

So far we have spoken only of the initial works of Lorentz, Poincaré and Einstein. It is to be hoped that this is adequate for a comparison of their relative significance. I should like to conclude by emphasizing that, naturally enough, work carried out after 1905 also contributed to the elaboration of special relativity. We could note papers by Einstein himself, as well as by Max Planck and, especially, Hermann Minkowski (his four-dimensional interpretation of the theory proved extremely fruitful).

II. The unsuccessful search for a medium for the propagation of light. Light phenomena *in vacuo* play a special part in the special theory of relativity. The speed of light *in vacuo* is the limiting speed with which signals can be transmitted, and it is right to say that the history of relativity theory begins with the discovery that the speed of light is finite.

As indicated in § 1.8, the theory of relativity proceeds from the consideration that the propagation of light (electromagnetic waves) requires no material medium; in other words, light can travel through vacuum. The idea entered physics with great difficulty, and it is associated with the special theory of relativity. Today it is part of the ABC of physics. The abandonment of the notion of a "luminiferous medium" under pressure of experimental facts is an extremely instructive page in the history of physics which is worth dwelling upon. We must, however, begin from afar and briefly recall the development of notions regarding the nature of light. In the early 17th century two points of view on the nature of light appeared, neither of which has lost its significance to this day. One, the "corpuscular", belongs to Newton; the other, the "wave", belongs to Huygens. Newton's initial premise can be readily appreciated: the success of his mechanics required a mechanical interpretation of light. Newton held that light represented the motion of special material particles called corpuscles. The basic properties of light — propagation in a straight line through a homogeneous medium, the laws of reflection and refraction — can be easily explained in terms of the corpuscular picture. In a homogeneous medium no forces act on the corpuscle, and it moves by inertia, i.e. in a straight line. Reflection occurs according to the law of elastic impact (like a billiard ball striking the cushion of the table). The angle of incidence equals the angle of reflection,

* It seems remarkable that there is not a single mention of Einstein's work on special relativity in any paper of Poincaré's although he died seven years after it was enunciated.

which is precisely the law of reflection of light from a stationary surface.

If a corpuscle at the interface of media *I* and *II* is subjected to forces normal to the interface in the direction of the denser medium, it changes the direction of motion. Indeed, let the velocity components of the corpuscle in medium *I* be V_n and V_t. The forces acting at the interface augment V_n, the direction of the velocity changes, and "refraction" occurs (Fig. S.1).

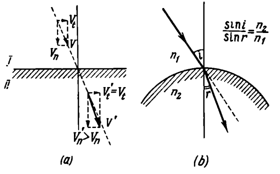


Fig. S.1. (*a*) The change in the direction of motion of a corpuscle in crossing the interface of two media *I* and *II* After crossing the interface $V_n' > V_n$, but $V_t' = V_t$, and the direction of the velocity changes (*b*) Refraction of light crossing the interface of two media.

We know from geometrical optics that when light passes from a medium with a refractive index n_1 into a medium with a refractive index n_2 refraction occurs. The connection between the angle of incidence i and the angle of refraction r is given by the Descartes-Snell law:

$$\frac{\sin i}{\sin r}=\frac{n_2}{n_1}.$$

Qualitatively, Newton's reasoning explains the refraction of light.

Newton also knew that light possessed properties which fitted with difficulty into his scheme (recall "Newton's rings" in optics, a typical interference phenomenon), but he devised ingenious explanations to maintain the corpuscular picture. At about the same time, Huygens put forward the idea of the wave nature of light. He proceeded from an analogy with sound waves, although he had no idea of the nature of light waves. He knew that light could travel where sound could not (if you place a bell under a transparent dome and evacuate the air you can see the hammer strike the bell but do not hear the sound). Huygens (and all physicists after him until Einstein) could not imagine vibrations propagating in the absence of any medium. So it was essential to introduce a

special medium through which light waves could travel. Huygens called it the *ether*. Thus appeared a concept the fallacy of which was revealed only by the theory of relativity.

Newton rejected the wave theory of propagation of light. He based his reasoning on the phenomenon of double refraction in crystals. Newton showed that if light propagates as waves then double refraction indicates a preferred direction of vibrations in the beam. But in Newton's time only longitudinal waves were known, which do not possess this property. So he rejected the wave theory, although he conceded that it was plausible. Newton also categorically rejected the ether *.

The corpuscular view of the nature of light was dominant for a hundred years after Newton's death. Its success was to a considerable degree due to his prestige. But a theory is good only as long as it does not contradict facts and explains them. The beginning of the 19th century brought with it the discovery of phenomena which offered convincing testimony of the wave nature of light. Light interference and diffraction were thoroughly investigated, and rectilinear propagation was satisfactorily explained on the basis of the wave theory. The discovery of polarization indicated that light waves were transverse.

Thus, the 19th century ushured in the triumph of the wave theory of light There seemed no doubt at all that light was a wave process. But 19th-century physicists could not imagine oscillations in the absence of some bodies or some medium. There had to be one. Its name — the ether — had already been coined by Huygens; it remained to establish its physical properties. Nineteenth-century physics was dominated by the ideas of mechanics, and it is hardly surprising that ether was endowed with the mechanical properties of a solid (transverse vibrations can propagate only through elastic solids). It was, of course, a queer solid indeed: it could not be sensed in motion, was invisible, could not be touched; but neither could it be endowed with other properties without coming into contradiction with observations.

But even setting aside the difficulties with the properties of the ether, another pertinent problem was that of a frame of reference with which the ether could be associated, i.e. the system in which it was at rest. Obviously, it would have to be a preferred system over all others, at least as far as optical phenomena were concerned. Here, nature itself seemed to offer a direct answer to the question, if the aberration of light was taken into account. The phenomenon consists in the following. If a ray of light is observed

* An idea of the debate between Newton and Huygens can be found in *The Evolution of Physics*, by A. Einstein and L. Infeld, New York, 1942.

from two reference systems moving relative to one another it will be seen at different angles to some direction common to the two systems (for instance, the direction of the relative velocity). If we observe the beams through a telescope, the visible direction coincides with the axis of the telescope. But why does the motion of the observer (telescope) affect the apparent direction of the incident light? This can be explained with the help of a simple example. Let a bead be falling vertically with a uniform velocity c. We want it to pass through a pipe of length l moving horizontally with a velocity V without hitting the walls. The way to achieve this is by keeping the bead on the pipe's axis BB''. For this, when the bead reaches point B', the lower end of the pipe B should arrive there at the same time. Obviously, the pipe should be tilted forward in the direction of the motion. The angle of inclination φ to the vertical is easily determined. Let the bead travel the distance $B''B' = l\cos\varphi$ in time τ. In the same time, the end B of the pipe must travel the distance $BB' = l\sin\varphi$. But $l\cos\varphi = c\tau$, and $l\sin\varphi = V\tau$, whence $\tan\varphi = V/c$ (Fig. S.2*a*).

In the corpuscular theory, the corpuscles play the part of the bead. Consequently, the telescope must be tilted forward in the direction of the motion. But the same reasoning holds in the wave theory of light. To keep the moving pipe from "indenting" the light wave front (the velocity of the pipe is V and that of light is c) it must be tilted at the same angle φ, such that $\tan\varphi = V/c$ (Fig. S.2*b*).

The aberration angle is defined as the change of the apparent angle at which the incident ray is observed in passing from one inertial frame of reference to another. Obviously, it is impossible to detect the aberration angle within one inertial frame, because the direction of the beam (toward a distant star) is always the same. Nevertheless the aberration of light was discovered by observing stars from the Earth, because the Earth moves in an ellipse, hence it is the same inertial reference system only over a limited time interval. Every six months the Earth reverses its direction of motion, and therefore the apparent sighting of the star should change.

The English astronomer James Bradley was looking for the parallactic shift of stars: the apparent path traced by a star over a year due to the change in the position of the observer. Fig. S.2*c* explains the appearance of the apparent parallactic shift of the North Star. In the course of a year it should describe a small ellipse occupying a very specific position relative to the orbit of the Earth.

In 1728, while trying to detect the parallactic shift, Bradley discovered the aberration of light; he found that stars lying near the pole of the ecliptic indeed describe an ellipse the major semi-

axis of which is equal to 41″. However, the ellipse did not lie as it should have for parallactic shift (Fig. S.2*c*).

Fig. S.2*d* shows how the aberration of light affects the apparent displacement of a star. Star *s*, lying perpendicular to the plane of the Earth's orbit, is observed from two diametrically opposite po-

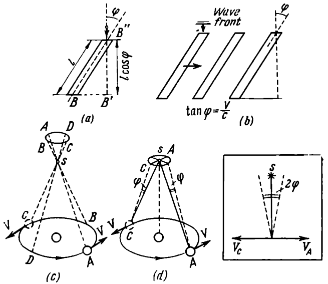


Fig. S.2. (*a*) A vertically falling bead must pass through a tube moving horizontally with velocity *V*. (*b*) A wave front must pass without distortion through a tube moving horizontally with the velocity *V* (*c*) The appearance of the parallactic shift of a star at the pole of the ecliptic due to the motion of the Earth Thanks to this effect, in the course of a year the star describes a small ellipse Attention must be given to the observer's location on the Earth; an observer at point *A*, for example, sees star *A* on the celestial sphere (the position of the star is also denoted by *A*) (*d*) The aberration of light and the Earth's motion in orbit also cause the star at the pole of the ecliptic to describe an ellipse in the course of the year However, the respective positions of the Earth and the star differ from those shown in (*c*). This is the difference between aberrational and parallactic shifts The diagram in the frame shows the appearance of the aberration angle when the Earth's motion reverses (six months later)

sitions *A* and *C*. Over six months the angle of the direction toward the star from these two points varies through 2φ. In Fig. S.2*d*, angle φ is the angle between the direction at which an observer would see the star *s* from a stationary Earth (which, of course, is impossible), and the apparent direction to the star *s*. Six months later the same angle will be in the opposite direction, and the difference between the apparent directions to the star *s* is 2φ. We shall use a simple calculation to evaluate angle φ. Light from the

North Star falls normal to the plane of the Earth's orbit. The motion of the Earth is perpendicular to the direction of the ray. The velocity of the Earth is 3×10^4 m/s, the velocity of light is 3×10^8 m/s. Hence, $\varphi = \arctan V/c = 20.5''$; $2\varphi = 41''$. This was the value Bradley obtained. He realized that he had failed to detect the parallax of a fixed star (it was discovered a hundred years later by Bessel) and instead discovered the aberration of light. Bradley explained it on the basis of the corpuscular theory, which we presented here. But the same result obtains for the wave theory as well. Thus, the explanation of aberration posed no difficulty for the wave theory. It did, however, involve an inevitable corollary regarding the "luminiferous ether". It had to be assumed that light travels through a medium at rest with respect to the heliocentric system, otherwise it would not impinge normal to the plane of the Earth's orbit.

It thus followed from the aberration of light that the ether was stationary in a heliocentric frame of reference (Newton's absolute system). The heliocentric system thus turns out to be a preferred, privileged one with respect to the propagation of light. Let us not forget that up to the mid-nineteenth century no one knew that light waves were electromagnetic waves of a specific frequency.

If we assume the existence of a luminiferous medium, its role differs in no way from that of any material medium transmitting vibrations. If the propagation speed of vibrations in a reference system in which the medium is at rest is v, then in any other system moving relative to the medium with the speed $\pm V$ the propagation velocity of the oscillations will be $c \pm V$. Under any wave theory the velocity of the waves is independent of the motion of the source, but it depends upon the motion of an observer relative to the medium the oscillation of which produces the waves: in our case the ether. Thus phenomena depend not only on the relative velocities of bodies, but also on their velocities relative to the medium

These assertions are best explained with the example of the Doppler effect for sound waves in air. Let the air be at rest in a system K where the velocity of sound is v; the source is moving relative to K (i.e. the air) with a velocity V, and the observer is at rest in the system (Fig. S.3*a*). Let us attach to the source a system K' (motion toward the observer) or K'' (motion from the observer). We can now reproduce the reasoning in § 3.4. Pulses are sent from the source at intervals T' or T'' ($2\pi/T'$ and $2\pi/T''$ are the proper frequencies ω_0). The receiver picks up two consecutive signals at intervals $T = T' - \frac{VT'}{v} = T'\left(1 - \frac{V}{v}\right)$ in the first case, and $T = T'' + \frac{VT''}{v} = T''\left(1 + \frac{V}{v}\right)$ in the second. Thus, if the source

is moving relative to the observer and the medium, when it approaches the observer the latter will observe an increase in the frequency $\omega = \omega_0 \big/ \left(1 - \frac{V}{v}\right)$, and when it recedes, a decrease $\omega = \omega_0 \big/ \left(1 + \frac{V}{v}\right)$.

If, now, the source is at rest relative to the medium, while the observer is receding from (or approaching) it, then the propagation velocity of the vibrations relative to the observer will be $v - V$ or $v + V$, respectively. Now the source is at rest in K, and the

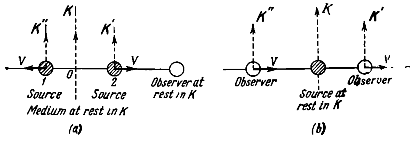


Fig. S.3. Illustration of the derivation of the Doppler effect formula for sound. (*a*) The observer and the air are at rest in the K system, and the source is moving through the medium with the velocity V. (*b*) The source of the sound and the air are at rest in the K system, and the observer is moving through the air with the velocity V

observers are connected with the systems K' and K''. Signals are emitted from the source at intervals T; they will reach the observer in K' at intervals $T' = T + \frac{VT}{v - V} = \left(\frac{1}{1 - \frac{V}{v}}\right) T$, and the observer in K'' at intervals $T'' = T - \frac{VT}{v + V} = \left(\frac{1}{1 + \frac{V}{v}}\right) T$. Only now $\omega_0 = 2\pi/T$, whence in receding we obtain a reduction in frequency $\omega = \omega_0 \left(1 - \frac{V}{v}\right)$, and in approach an increase, $\omega = \omega_0 \left(1 + \frac{V}{v}\right)$.

We find that the equations are different for the same relative velocity of the source and the observer. Similarly to what was done in § 3.4 for the case of wave radiation at an angle to the direction of propagation, we obtain $\omega = \omega_0 \left(1 - \frac{V}{v} \cos\theta\right)$ and $\omega = \dfrac{\omega_0}{1 - \frac{V}{v}\cos\theta}$.

This example readily demonstrates that the ether assumption violates the relativity principle. Of course, the relativity principle

is valid in the presence of a medium, but to make the conditions of the experiment identical, the velocity of the reference system relative to the medium must be the same. In other words, every reference system must be associated with a definite medium. Thus, for the relativity principle to hold the ether would have to be entrained, or partially entrained by the reference system (the ether drag). This is a strange assumption, but nevertheless it was developed, though in a different connection.

If the hypothesis of the stationary ether were correct it could explain other optical phenomena. From this point of view, Fizeau's experiment (1851) led to some very mysterious results We already described the experiment in § 3.5, but let us take a look at it from the point of view of a physicist of the 19th century. Let the water be at rest in the system K', which is moving together with the water with velocity V relative to the laboratory. Let us also assume that the laboratory can be considered the preferred reference frame in which the ether is at rest. Light propagates through the ether The substance changes its phase velocity, but the velocity of the substance does not matter. Consequently, the velocity of light in the laboratory system, v, is simply equal to the velocity of light, v', in the stationary water. Of course, $v' = c/n$. Assume for a moment that the ether is "dragged" together with the water. Then, naturally, the velocity of light, v', would be compounded with the velocity of the ether, i.e. the water, and we would get $v = v' + V$. An ether endowed with the strange property of "partial" drag would give $v = v' \pm kV$, the sign depending upon the relative direction of motion of the light and the medium. Fizeau's experiment, which was subsequently repeatedly confirmed, gave the result $v = v' \pm \left(1 - \frac{1}{n^2}\right) V$. In any event, the stationary ether hypothesis contradicted the results of Fizeau's experiment.

The battle for recognition of the ether did not end with this, of course, but before describing further attempts to detect the ether it is useful to dwell on the relationship between physical experiment and theory. Cognition of nature involves the search for laws that correctly reflect existing relationships or, philosophically speaking, the search for objective laws of the objective world existing independently of us. Nature is studied by people, and they invariably introduce their subjective notions and sensations, to say nothing of inevitable errors. Consequently, the laws of nature must be verified. What is the criterion of correctness of a physical law? The correctness of physical laws is revealed in practical activity. The guarantee that our knowledge agrees with the laws of nature lies in their verification by different people in different places, and repeated verification by one person The most important means of verifying physical laws and discovering the laws of nature is by

artificially creating the necessary conditions, i.e. by staging a physical experiment.

But physics cannot be restricted to experimental results. Physics is impossible without theories that can be used to systematize and explain various natural phenomena on the basis of a small number of fundamental laws. In turn, physical theory — and today it is a science in its own right, theoretical physics — is closely linked with mathematics. When experimental material accumulates, a theory appears to explain a specific set of phenomena. Can an experiment or series of experiments vindicate or refute a theory? We are not, of course, speaking of erroneous experiments, which may always occur; ultimately their erroneousness is sure to be shown. Sometimes in developing a theory the limits of its application are apparent, and one should keep within them. With these reservations, the following can be asserted: if a single correct experiment carried out within the limits of applicability of a given theory contradicts that theory, the theory must be assumed wrong. As for "proof" of a law or theory through comparison of obtained conclusions with experimental data, no amount of experiments agreeing with a theory can be regarded as the ultimate proof. A theory exists and is considered correct as long as its conclusions do not contradict some new experiment within the domain of its applicability. The situation is analogous to one of the rules of mathematics: the validity of special cases is not proof of the validity of a general theorem, but one example to the contrary refutes it.

Physics is essentially an experimental science. The process of cognition of nature, of which physics is a part, is continuous and unlimited. The question of establishing the ultimate truth is, rather, a philosophical issue than one of some specific science. Individual physical experiments reveal various specific laws and regularities or, from the philosophical point of view, relative truths. Although the regularities contain elements of absolute truth, they do not provide exhaustive knowledge. In the final analysis, every theory is either restricted or erroneous. At each stage a theory's validity is determined by the absence of experimental data contradicting it, while its value lies in its ability to explain and predict observable phenomena. These remarks, interesting in themselves, need not have been cited if they were not useful in setting forth the history of the ether. We shall rely mainly on "negative" experiments pointing to the fallacy of various premises postulated in an attempt to salvage or discover the ether.

Let us return to the assumption that the ether is stationary in a heliocentric system (which follows from the observed aberration of light). If the ether is stationary in a heliocentric system, the Earth in its motion around the Sun should experience an "ether wind". It is not hard to conceive an experiment to discover motion

relative to the ether. It is shown graphically in Fig. S.4. Two photoelectric cells are sitting on an optical bench parallel to the Earth's velocity in orbit. Halfway between them is a light source, I, which emits flashes of light. Light travels through the stationary ether with the velocity c. But cell C_1 is moving toward the light beam, and cell C_2 is receding from it. Light travels toward C_1 with the velocity $c+V$, and toward C_2 with the velocity $c-V$. Thus,

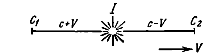


Fig. S.4. A light source, I, and two photoelectric cells, C_1 and C_2, are mounted on an optical bench. The bench is positioned parallel to the direction of the velocity of the Earth in its orbit. If light propagates in the stationary "ether" it travels from the source I toward cell C_1 with the velocity $c+V$, and toward cell C_2 with the velocity $c-V$.

C_1 will register the arrival of the light beam before C_2, the time interval being

$$\Delta t=\frac{l}{c-V}-\frac{l}{c+V}=2l\frac{V}{c^2}\frac{l}{1-\frac{V^2}{c^2}}. \qquad \text{(S.II.1)}$$

Before considering the possibility of carrying out such an experiment, note the following. In speaking of the velocity of light, the ether assumption reduces solely to the assertion that there exists a single reference frame in which the velocity of light *in vacuo* is c. It is the one and only preferred system. In all other frames moving relative to it with velocity V, the velocity of light *in vacuo* is given by the classical rule of addition of velocities, $c'=c\pm V$. Our reasoning applies to a reference frame K connected with the ether (or simply with the frame in which the velocity of light *in vacuo* is c). It is assumed here that the signals received by the photoelectric cells are registered in the K frame. But if we apply the same reasoning to a frame K' in which the optical bench is at rest we obtain exactly the same result. Simply, in that frame the velocity of light *in vacuo* differs from that in K: $c-V$ to the right, and $c+V$ to the left (Fig. S.4). In the K frame, we recall, $c-V$ and $c+V$ would be simply the velocities with which light reaches cells C_1 and C_2 *. It is not surprising that the results are the same in K and K', since in classical mechanics the timing of events is absolute.

Thus, if we could measure the time difference Δt (S.II.1) we would thereby only prove the difference in the speed of light in

* It is worth remembering that according to the STR the velocity of light in the K' frame is the same as in K, and equal to c.

frames K and K'. That, of course, would be indirect proof of the ether (see further on).

As to the order of magnitude of Δt in such an experiment, if we put, for example, $l \approx 100$ m, and $V \approx 3 \times 10^4$ m/s (the speed of the Earth on its orbit), then $\Delta t \approx 10^{-10}$ s. Even modern hardware cannot measure such a minute time interval.

In 1878, Maxwell suggested an experiment employing the phenomenon of interference of light to detect the motion of the Earth relative to the ether, at rest in a heliocentric system (if light actually propagated through it). Maxwell thought the required accuracy of measurement to be unattainable, but three years later Michelson built an interferometer capable of detecting the motion of the experimental set-up relative to the stationary ether.

Michelson's experiment, carried out in 1881 *, was as follows (Fig. S.5). A beam of light from a source, I, is directed on a half-silvered glass plate, P. Half the incident light is reflected, the other half passes through the glass. Michelson's instrument (today it is called *Michelson's interferometer*) had two mirrors, S_1 and S_2, located as shown in the drawing, at distances L_1 and L_2 from the half-silvered glass P. All the parts of the interferometer were firmly secured to a heavy block of stone floating on a disc of wood in a tank of mercury so that the whole system could be turned smoothly.

At the plate P the beam is split in two: beam *1* travelling to mirror S_1, and beam *2* to mirror S_2. Each beam reaches its mirror and returns to plate P. As the plate is semi-transparent, a portion of the light from both beams travels in direction *3*. Since each of the beams *1* and *2* is a portion of the initial beam, the two rays *1* and *2* travelling in the direction *3* are coherent and can interfere.

Let us determine the time it takes the beam of light to travel from P to S_2 and back. The interferometer is at rest in a system which is travelling with a velocity V relative to the ether. The distance between P and S_2 is L_2, the speed of light to the right is $c - V$, and to the left, $c + V$. Hence, the required time is

$$t_2 = \frac{L_2}{c - V} + \frac{L_2}{c + V} = \frac{2L_2}{c} \frac{1}{1 - \frac{V^2}{c^2}} \tag{S.II.2}$$

Let us now find the time $\tilde{t}_1$ it takes beam *1* to travel from P to mirror S_1. In the time $\tilde{t}_1$ mirror S_1 travels the distance $V\tilde{t}_1$, and the light travels a distance $c\tilde{t}_1$ along the hypotenuse of triangle $PS_1'P''$. From this right-angle triangle it follows that $(c\tilde{t}_1)^2 = L_1^2 + (V\tilde{t}_1)^2$, whence

$$\tilde{t}_1 = \frac{L_1}{\sqrt{c^2 - V^2}} = \frac{L_1}{c} \frac{1}{\sqrt{1 - \mathrm{B}^2}}, \quad \mathrm{B} = \frac{V}{c}.$$

* See A. A. Michelson and I. W. Morley, *Phys. Mag.* (5) 24, 449 (1887) — *Ed.*

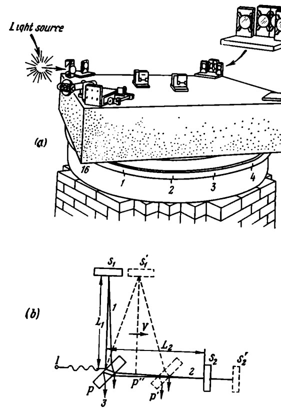


Fig. S.5. (*a*) Michelson's interferometer (*b*) Diagrammatic representation of Michelson's experiment. A beam of light from the source *I* is split on the semitransparent plate *P* into two beams, *1* and *2*, travelling along and perpendicular to the direction of the Earth's motion in its orbit. The velocity of the Earth's orbital motion is indicated by the arrow and the letter *V*. Beams *1* and *2* reflect from mirrors S_1 and S_2, respectively, and return to plate *P*. After reflecting and refracting, two beams travel in the direction *3*. That is the direction in which the interference pattern is observed. The decisive element of the experiment consists in turning the whole apparatus through 90°; beam *1* now travels in the direction of the Earth's motion, and beam *2* normal to that direction. If light propagated through the stationary ether the difference in the paths of beams *1* and *2* would change and the interference pattern observed in direction *3* would change (the interference fringes would shift). The experiment, however, discovered no shift in the interference fringes.

The total time it takes the light beam to travel from P to S_1 and back is double that amount, so that

$$t_1 = 2\tilde{t}_1 = \frac{2L_1}{c}\frac{1}{\sqrt{1-B^2}}. \qquad \text{(S.II.3)}$$

The difference between times t_1 and t_2 is due not only to the difference in the lengths of the arms of the interferometer but also to the motion of the apparatus.

One of the arms of the interferometer is aligned parallel to the motion of the Earth along its orbit (which is known from astronomical observations). But we live on Earth which, according to the assumption, is moving relative to the ether, hence the experiment is inevitably conducted in an "ether wind". Hence, too, it is impossible to compare the interference pattern "without the ether wind" and "in the ether wind". By aligning, for example, arm PS_1 in the direction of the velocity V, we obtain a certain interference pattern: an alternation of light and dark fringes *, depending on the difference in the propagation of beams *1* and *2*:

$$\Delta t = t_2 - t_1 = \frac{2}{c}\left(\frac{L_2}{1-B^2} - \frac{L_1}{\sqrt{1-B^2}}\right) =$$
$$= \frac{2}{c\sqrt{1-B^2}}\left(\frac{L_2}{\sqrt{1-B^2}} - L_1\right). \qquad \text{(S.II.4)}$$

This time the difference depends on both L_1 and L_2 and the velocity V. If the condition $L_1 = L_2$ could be guaranteed the difference would depend on $L_1 = L_2 = L$ in this way:

$$(\Delta t)_{L_1=L_2} = \frac{2L}{c}\left(\frac{1}{1-B^2} - \frac{1}{1-B^2}\right) \qquad \text{(S.II.5)}$$

When L is known, the interference fringes observed in the apparatus are uniquely determined by its motion. But, first, it is simpler not to worry about assuring the condition $L_1 = L_2$, and secondly, it is more convenient to observe the change in the interference pattern, that is, the displacement of the fringes. For that the whole set-up is rotated through 90°. The arms of the interferometer change places. We then obtain

$$t_1' = \frac{2L_1}{c}\frac{1}{1-B^2}, \quad t_2' = \frac{2L_2}{c}\frac{1}{\sqrt{1-B^2}}, \qquad \text{(S.II.6)}$$

$$\Delta t' = t_2' - t_1' = \frac{2}{c}\left(\frac{L_2}{\sqrt{1-B^2}} - \frac{L_1}{1-B^2}\right) = \frac{2}{c\sqrt{1-B^2}}\left(L_2 - \frac{L_1}{1-B^2}\right).$$

* On Michelson's interferometer see, for example, the already quoted *Optics* by G. Landsberg.

This means that in the rotation of the apparatus the time difference changes as follows·

$$\Delta t^{*} = \Delta t' - \Delta t = \frac{2(L_1 + L_2)}{c\sqrt{1 - B^2}}\left(1 - \frac{L}{\sqrt{1 - B^2}}\right) \approx -\frac{L_1 + L_2}{c} B^2, \quad \text{(S.II.7)}$$

and the fringes shift.

If we want to obtain a difference in the distance travelled by the beams of the order λ, Δt^{*} must be of the order of the vibration period, T. But $T = \lambda/c$. For the motion of the Earth in orbit $B \approx 10^{-4}$; for visible light $\lambda \approx 5 \times 10^{-5}$ cm, whence the total length of the interferometer arms $L_1 + L_2 \approx 50$ metres. Such a path for the light beam can be obtained by repeated reflection.

Michelson could have detected an "ether wind" blowing at 10^4 m/s. Together with all other physicists, he had not the slightest doubt that such an "ether wind" would certainly be felt. But there was none. The experiment was repeated many times, with greater and greater precision: today an "ether wind" blowing at 30 m/s could be detected, but Michelson's result, or his null result, as it is also called, stands fast. There is no doubt that it is correct.

However, it does not necessarily follow from Michelson's experiment that there is no ether. Its results can be explained by endowing the ether with certain properties. The *coup de grace* against the ether required other observations. But let us first draw some conclusions from Michelson's experiment, without linking it with the search for the "ether wind". The experiment showed that a rotation of the interferometer on Earth does not cause a shift in the interference fringes. In principle, however, such a shift could be associated with the difference in the speed of light along the two directions in the frame of reference of the interferometer.

Thus, regardless of whether the ether exists or not, Michelson's experiment shows that the speed of light over a closed path measured on Earth is the same in all directions, i.e. it is isotropic. As the Earth moves around the Sun along a closed curve, and can therefore be regarded as an inertial frame of reference only over a short time interval, our measurements are actually carried out in many inertial reference frames. It thus follows from the experiment that the velocity of light along a closed path is isotropic in any inertial frame of reference

A somewhat modified form of Michelson's experiment, staged in 1932 by Kennedy and Thorndike, offers confirmation of Einstein's main postulate. From equation (S.II.7) it is apparent that Δt also depends upon c. If the absolute velocity of light *in vacuo* were different in different inertial frames of reference, then a shift in the fringes would be observed in passing from one IFR to another. The interference pattern was on several occasions observed continuously for periods ranging from eight days to one month (with a three-

month interval). During that time the apparatus changed many inertial frames of reference. It is so sensitive that a variation of 2 m/s for c could be detected. But nothing happened. It thus follows from Michelson's experiment that the speed of light has the same value (travelling there and back) in all directions within a given IFR; moreover, it has the same value *in vacuo* in all inertial frames of reference.

At the risk of interrupting our discourse, it is worth presenting the results of Michelson's experiment in terms of the special theory of relativity. If the experiment is staged in an inertial frame of reference its result is obvious. In every such system the velocity of light *in vacuo* is isotropic and the interference pattern observed in the experiment is due solely to geometric differences in path. In short, in every IFR in which the interferometer is at rest we obtain exactly the same picture as we would have in a *classical consideration* in the preferred reference system in which the ether is stationary.

The need for a more complex interpretation of the result arises when the experiment is considered in an inertial reference frame relative to which the interferometer is in motion. Let the interferometer be at rest in K' and the experiment be staged in K; for the sake of simplicity we assume that in K' $L_{10} = L_{20}$. We have added the subscript 0 to stress that we are dealing with proper lengths. Obviously, equations (S.II.4) and (S.II.6) remain valid, but $L_2 = L_{20}\sqrt{1 - B^2}$, and $L_1 = L_{10}$. It is apparent then that Δt, determined according to (S.II.6), vanishes. Naturally, a rotation of the apparatus yields no effect. Thus, in special relativity the null result of Michelson's experiment is explained by the relativity of the length of the measuring rods (which is most directly linked with the invariance of the velocity of light, § 2.3).

To explain the result of Michelson's experiment and salvage the ether, Lorentz and Fitzgerald assumed that in motion relative to the stationary ether all bodies contract by the factor $\sqrt{1 - B^2}$ in the direction of motion (the "Lorentz contraction"). It is apparent from the foregoing reasoning that such an explanation is possible. But it is essential to emphasize the difference in the relativity of the length of measuring rulers in special relativity and in the Lorentz contraction. In the STR the contraction is a consequence of measurements made in the relative motion of reference frames. With Lorentz-Fitzgerald it is the consequence of motion relative to the ether, which retains the preferred reference system. The fallacy of the Lorentz contraction is revealed in a modification of Michelson's experiment. Michelson's interferometer had arms of the same length; Kennedy and Thorndike's had arms of different length. If the interferometer arms are different, a shift

in the interference fringes should be observed when the interferometer's velocity relative to the ether changes. But on Earth an interferometer participates in three motions relative to the ether: the Earth's motion relative to the Sun, its rotation, and finally, the motion of the Sun. The resultant velocity varies by a certain value every 12 hours (and six months). These variations should result in a shift in the interference fringes. Indeed, if Lorentz's hypothesis is correct, then $L_2 = L_{20}\sqrt{(1-B^2)}$, and from (S.II.4) we get

$$\Delta t = \frac{2}{c\sqrt{1-B^2}}(L_{20}-L_{10}). \qquad \text{(S.II.8)}$$

Without rotating the apparatus, let us see how Δt varies with a variation of the velocity B relative to the ether by ΔB. Differentiating (S.II.8) with respect to B, we obtain

$$\frac{\Delta t}{\Delta B} = \frac{2}{c}(L_{20}-L_{10})\,\Delta\left(\frac{1}{\sqrt{1-B^2}}\right) = \frac{L_{20}-L_{10}}{c}\,\Delta B^2. \qquad \text{(S.II.9)}$$

The latter equation is written to the accuracy of B^2, but prolonged observations of the interference pattern revealed no variations.

Another way to square the ether with the results of Michelson's experiment is to assume that the ether is "dragged along" by moving bodies. But as we have seen, the aberration of light "agrees" only with a "stationary ether in a heliocentric system". An experiment specially staged by Fizeau (see § 3.5, where it is explained in terms of special relativity) to determine the ether "drag" led to the conclusion that there was a "partial drag".

Of course, the mentioned experiments and observations far from exhaust the attempts to establish the properties of the ether. But it is already obvious that the ether would have to be endowed with extremely contradictory properties. But the ether's most significant "contribution" to physics would probably have been the rejection of the relativity principle in electrodynamics.

In 1905, Einstein's paper, *On the Electrodynamics of Moving Bodies*, appeared. It virtually set forth the whole of the special theory of relativity, which not only offered a natural explanation of the result of Michelson's experiment but also correctly interpreted all known mechanical, electrodynamic and optical phenomena. From the outset it extended the relativity principle to all physics and unambiguously asserted the equality of all inertial frames of reference *in vacuo*, thereby making the ether redundant. Not a single experimental fact contradicts special relativity.

In his initial experiments Michelson could have discovered a variation in the speed of light in a change of its direction relative to the motion of the Earth down to 0.15 m/s; later experimenters could have detected a variation of 0.015 m/s; with lasers the de-

tectable change in speed is a mere 3×10^{-5} m/s. Yet no variation in the speed of light relative to a moving observer has ever been detected.

That the speed of light *in vacuo* does not depend on the motion of the source has been repeatedly verified. One experiment was carried out with an extraterrestrial source, the Sun (A. M. Bonch-Bruevich, 1956). If the velocity of light depends on the motion of the source, then by measuring the velocity of light emitted from two opposite points of the equator it should be possible to detect the difference between those velocities. No such difference was detected. A laboratory experiment was conducted in which the flight of gamma-quanta over a certain distance was compared; gamma-quanta emitted by a stationary and a moving source (radioactive nuclei) were studied, and again the independence of the velocity of light was confirmed.

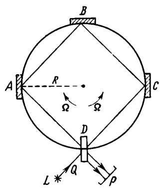


Fig. S.6. Diagrammatic representation of the Sagnac-Garress experiment

It can be confidently declared that, despite the enormous increase in the accuracy of experiments, there is no indication of the existence of a preferred reference system, or of any difference in the velocity of light *in vacuo* in different inertial frames of reference, or of any manifestation of the ether

In conclusion we should note that *accelerated* motion of a reference system relative to an inertial frame of reference can, of course, be detected. A mechanical experiment of this kind — Foucault's experiment — was described in § 1.5. There are also optical variants of the experiment, which we shall mention to round out the picture. We shall describe Garress's experiment (1912), subsequently repeated by Sagnac. Three mirrors, *A*, *B*, *C*, and a semi-transparent plate *D* are mounted on a turntable (Fig. S.6) together with a light source *L* and photographic plate *P*. A beam of light *Q* is split at plate *D* into two beams, *DABCDP* and *DCBADP*, travelling along the path *ABCD* in opposite directions. If the system is at rest, interference of the two beams travelling from *L* and split at *D* occurs on the plate. When the turntable rotates the interference fringes should displace owing to the change in path lengths.

Disregarding the deviation of a geocentric system from an inertial system, let us examine events in the reference system of the Earth. For the sake of simplicity we shall assume that there

are many mirrors and the path of the beams is virtually a circle. Then the speed with which the light catches up with the turntable when it travels in the direction of rotation is $c - V = c - \Omega R$ (where R is the radius of the turntable and Ω is the angular velocity of rotation); for the beam travelling in the opposite direction the velocity is $c + V = c + \Omega R$. The time it takes the beam to travel around the circumference is $\tau_1 = 2\pi R/(c - V)$ in the first case, and $\tau_2 = 2\pi R/(c + V)$, in the second. The difference between the two times is

$$\Delta\tau = \tau_1 - \tau_2 = 2\pi R\left(\frac{1}{c-V} - \frac{1}{c+V}\right) = \frac{4\pi RV}{c^2}\,\frac{1}{1-\frac{V^2}{c^2}} \approx \frac{4S\Omega}{c^2},$$

where S is the area of the turntable. Sagnac observed a shift in the fringes that agreed nicely with this formula. The shift can be used to determine the angular velocity Ω.

If the Earth is used as the turntable its angular velocity can be determined. This experiment was carried out in 1925 by Michelson and Heyl. The angular velocity corresponded to the component of the angular velocity of rotation of the Earth along a plumb line at the point of observation. For the experiment two kilometres of pipes were laid and a second circuit was built to determine the zero point of displacement of the fringes. Michelson's result was 0.230 ± 0.005, the theoretical figure being 0.236. Excellent agreement!

Thus, unlike uniform translational motion of the Earth, its rotation can be detected by various physical experiments.

III. Was Michelson's experiment "decisive" for the creation of the special theory of relativity? Michelson's experiment is given great prominence in virtually all books on the history of special relativity. Most authors assume, one way or another, that special relativity was an upshot of attempts to explain Michelson's experiment, which is the theory's principal experimental basis

That is the place assigned to Michelson's experiment in the only book in Russian devoted to the experimental foundations of the STR, written by S. I. Vavilov in 1928: "The story of [Michelson's experiment] is set forth here in fairly great detail because the basic postulates of the relativity theory were formulated on its basis." And in his foreword Vavilov wrote, "After reading this book the reader will understand why it is adorned with a picture of Michelson."

This claim is repeated in virtually all our textbooks dealing with the history of special relativity. In Y. B. Rumer and M. S. Ryvkin's *The Theory of Relativity* (Moscow, 1960) we find: "Unlike all preceding investigators, Einstein saw the negative result of Michelson's experiment as ...". Foreign authors toe the same line on

this score. To cite but one example, Laue, in a book written in 1911, states that Michelson's experiment "became, as it were, the fundamental experiment for the relativity theory."

Besides textbooks and popular expositions there are, of course, books on special relativity written by Einstein himself. In his book on the special and general theories of relativity, which he subtitled, *A Comprehensible Exposition* (gemeinverständlich), Michelson's experiment is mentioned in § 16: "The Special Theory of Relativity and Experiment". However, it is not clear from the text whether there was any direct connection between Michelson's experiment and the enunciation of the theory. It is not mentioned at all. Nothing in other writings and statements of Einstein indicated any contradiction with the accepted view that Michelson's experiment was actually the point of departure for special relativity. Students were taught so, and schoolchildren are told as much today. It was therefore most surprising to read in an article by R. Shankland, published in 1963, the following excerpt from his interview with Einstein dating back to 1950:

"When I asked him how he had learned of the Michelson-Morley experiment, he told me that he had become aware of it through writings of H. A. Lorentz, but only after 1905 had it come to his attention! 'Otherwise', he said, 'I would have mentioned it in my paper!'" Indeed, Einstein's 1905 paper contains no mention of Michelson's experiment or references to Lorentz's papers.

We know that Lorentz and Poincaré came very close to postulating special relativity, but in fact it was enunciated by Einstein, and virtually in that single work in 1905 Thus, if we speak of the decisive significance of Michelson's experiment for developing special relativity, we should determine its influence on Einstein's work. And from Shankland's article we learn that Einstein first heard of Michelson's experiment only after the creation of the special theory of relativity.

Why should the question of the role of Michelson's experiment in Einstein's efforts to formulate special relativity be of such concern to the teacher? It is intriguing, of course, as a curious fact, but hardly worth devoting a whole paragraph to it. But it is not just a matter of curiosity. It is inevitable that the history of the development of a theory should be reflected in teaching. It is hard, and in some cases impossible, to sidestep the history of an issue. With time, of course, the presentation of a discipline becomes logically streamlined (teaching is never in vain!), but the real history of a theory's development does not always reflect the logic of its elaboration. Nature does not necessarily reveal its secrets in a sequence most convenient for their interpretation. In the initial period after the enunciation of a theory its academic presentation follows rather in the historical wake of its elaboration than accord-

ing to the logical scheme which can be constructed after the theory has been completed.

Thus, looking from this aspect at the academic presentation of special relativity, we can see, that, if Shankland understood Einstein correctly, the teaching of the theory did not even follow the steps in its creation. A curious situation!

The question is not of denying the role of Michelson's experiment. Its history and implementation cannot but arouse admiration. Michelson's experiment occupies an outstanding place in the history of natural science. And yet it played a rather unfortunate part in the evolution of the traditional scheme of describing special relativity.

When placed at the basis of instruction in the STR, Michelson's experiment inevitably introduces the ether. It is impossible to explain its meaning without speaking of the difficult search for a material medium through which light propagates (see Supplement II). But today we all know only too well that no such medium is necessary, and from the methodological point of view that is where we should begin. There is absolutely no need to go back to the intellectual atmosphere of the later nineteenth century.

Indeed, the ether played a prominent part in the physical views of the 19th century. It was, in fact, the ether concept which suggested to Maxwell the idea of the experiment ultimately carried out by Michelson and Morley. But erroneous notions, which may have played their part at a certain stage in the development of science, are ultimately discarded. When Galileo enunciated his inertia principle he immediately discarded Aristotle's doctrine that motion had to be continuously supported. When the transformation of mechanical energy into heat was discovered the phlogiston concept had to be discarded. The theory of relativity began with the rejection of absolute motion and the ether. But nowadays no one brings up Aristotle's doctrines in expounding mechanics, no one recalls phlogiston in lecturing on heat; why then should the ether be kept in describing special relativity at school and college? The introduction of the ether in modern expositions of special relativity is, to say the least, strange. First one has to explain at length the reason for introducing the ether, and then conclude by declaring that it does not, after all, exist. Can such an approach be called methodical?

It is sometimes said that there is no getting away from the ether: it had to be postulated by analogy with the propagation of sound or waves on water. One should only be sure to explain at the right time that there is a substantial difference between the way electromagnetic and gravitational waves travel, on the one hand, and elastic waves, on the other, and that matter is not required for the propagation of electromagnetic waves. Light can propagate in the

absence of matter in the conventional sense of the word (possessing rest mass). So there is no reason for the ether after all!

To this one could reply, "True, it is better not to introduce the 'ether', which doesn't exist anyhow, but special relativity is a complicated thing. The road to it was long and difficult, and it lay through the ether hypothesis, which was discarded when the need for it passed. But it was a natural theory which appeared in the search for a correct solution, it reflects the logic of human thinking, and there is no harm in setting it forth." Everything in this reasoning is correct except one thing. Einstein arrived at special relativity not via the ether (nor Michelson's experiment), but along a simpler and clearer road. And if one speaks of the logic of human thinking it is worth taking a closer look at Einstein's reasoning.

What role did Michelson's experiment play in Einstein's work? We have already cited Einstein's statement as quoted by Shankland. But Shankland's article appeared after Einstein's death and had not been, so to say, "authorized". Not long ago a letter written by Einstein and containing a direct answer to the question was discovered in the Einstein archive at Princeton, and it removes all doubts. The story of the letter is as follows. On February 2, 1954, a year before Einstein's death, a certain Mr Davenport wrote him, saying that he was looking for evidence that Michelson had "influenced your thinking and perhaps helped you to work out your theory of relativity." Not being a scientist, he asked Einstein for "a brief statement in non-technical terms, indicating how Michelson helped to pave the way, if he did, for your theory."

Einstein replied almost immediately, on February 9, 1954. It is his last statement on the question. It seems obvious that he had reflected about it before (after all, he had spoken of it with Shankland). The letter is clear and unequivocal. Here it is

"Dear Mr Davenport:

"Before Michelson's work it was already known that within the limits of the precision of the experiments there was no influence of the state of motion of the coordinate system on the phenomena, resp. their laws H. A. Lorentz has shown that this can be understood on the basis of his formulation of Maxwell's theory for all cases where the second power of the velocity of the system could be neglected (effects of the first order)

"According to the status of the theory, it was, however, natural to expect that this independence would not hold for effects of second and higher orders. To have shown that such expected effect of the second order was *de facto* absent in one decisive case was Michelson's greatest merit. This work of Michelson, equally great through the bold and clear formulation of the problem as through the ingenious way by which he reached the very great required precision of measurement, is his immortal contribution to scientific

knowledge. This contribution was a new strong argument for the non-existence of 'absolute motion', resp. the principle of special relativity which, since Newton, was never doubted in mechanics but *seemed* incompatible with electro-dynamics.

"In my own development Michelson's result had not had a considerable influence. I even do not remember if I knew of it at all when I wrote my first paper on the subject (1905). The explanation is that I was, for general reasons, firmly convinced how this could be reconciled with our knowledge of electro-dynamics. One can therefore understand why in my personal struggle Michelson's experiment played no role or at least no decisive role.

"You have my permission to quote this letter. I am also willing to give you further explanations if required

"Sincerely yours,
Albert Einstein"

The letter is clear, and there is nothing to add. True, it is contradicted by one pronouncement of Einstein's known from B. Jaffe's book. Einstein met Michelson but once, in 1931, in Pasadena. In a short speech, before those present, he said addressing Michelson, "Through your marvelous experimental work [you] paved the way for the development of the theory of relativity." Jaffe's book, from which these words are quoted, does not give Einstein's speech in full. From an account of it in German it follows that the book omits a whole sentence and that Einstein spoke of the "road" to the general theory of relativity. Thus, there is not the slightest hint on Einstein's part that Michelson's experiment was in any way decisive, although Einstein invariably stresses its beauty and fundamental contribution to science. In a letter to Jaffe [26] he writes that the experiment "strengthened my conviction concerning the validity of the principle of the special theory of relativity." That is probably the most correct assessment of the significance of Michelson's experiment for Einstein's work. The experiment's significance in the history of physics was quite different, perhaps even "decisive". The experiment's "null" result quite obviously dominated the work of Lorentz and many others. But that was not the road along which special relativity evolved. Paradoxically, Lorentz, who discovered the famous "transformations" that bear his name and which embody the very essence of special relativity, remained a long way off from enunciating the STR.

Finally, of interest is the question why academic instruction has been so persistent in following the "Lorentz way"? One fortuitous consideration which probably played no mean part ought to be mentioned. Here is a small quotation from the book by H. Bondi [21]:

"What has bedevilled this issue in textbooks is the undue prominence given to the Michaelson-Morley experiment .... Einstein said that at the time he wrote his basic paper on relativity (1905) he had never heard of the experiment. Later on when it was decided to reprint various essays on relativity it was decided by the publishers (with the advice of somebody) to start in the middle of one of Lorentz's essays. The first part that was included happened to be the Michaelson-Morley experiment. For this reason since then everybody, or nearly everybody, has felt obliged to start in the same way. And what a complicated start it is!"

Indeed, if we take Lorentz's book *Versuch einer Theorie der elektrischen und optischen Erscheinungen in bewegten Körpern* (Leiden, 1895), we see in §§ 89-92 a description of Michelson's experiment.

As for the exclamation, "And what a complicated start it is!", far from everyone agrees with it. But the question of how most reasonably to explain special relativity to the student is not an idle one: today special relativity is not only a part of the college course in general physics, it is also present in the high school curriculum. Instruction should, doubtlessly, be based on the ideas of modern physics and not include outdated notions of the past.

Einstein's elaboration of special relativity began with his rejection of the "luminiferous ether", and in that sense Michelson's experiment was certainly not "decisive". Einstein's reasoning is sufficiently simple and logical, and there is every reason to use it in expounding the special theory of relativity.

IV. Why shouldn't the mass-velocity dependence, or the relativistic mass, be introduced? Textbooks on special relativity (especially the older ones) often introduce the "relativistic mass",

$$m_{rel} = m\gamma = m/\sqrt{1-\beta^2}, \quad \beta = v/c,$$

which, by definition, depends upon the velocity, and try to give it independent meaning.

Whether the relativistic mass should be introduced or not is a purely methodological issue. Whether m_{rel} should simply be considered an abbreviated notation poses no problem at all. But the physical interpretation of relativistic mechanics is an entirely different question. Here misunderstandings and vague interpretations often arise. On this is our discourse.

It is not hard to understand how the temptation to introduce the relativistic mass appeared. One need but juxtapose the Newtonian and relativistic equations of motion (5.37a) and (5.37b),

$$\frac{d}{dt}(m\boldsymbol{v}) = \boldsymbol{F}, \quad \text{(a)} \quad \Big| \quad \frac{d}{dt}(m\gamma\boldsymbol{v}) = \boldsymbol{F}, \quad \text{(b)}$$

for the conclusion to suggest itself that the "only difference" between them is that in the relativistic equation the mass depends upon the velocity. The expression $m\gamma$ is then taken from under the derivative sign and declared the relativistic mass, with an independent meaning attached to it. There are many objections against such an interpretation; they will be set forth later on. On the other hand, emphasis will be laid on the advantages of employing the invariant rest mass.

A reasonable relativistic interpretation of mass should, like all relativistic mechanics, in the final analysis, rely on four-dimensional concepts. Although in many cases academic instruction is so concise that the introduction of four-dimensional concepts is impossible, we cannot forget that the construction of relativistic mechanics (Chapter 5) inevitably requires the introduction of a four-dimensional world. If we are forced to restrict ourselves to three-dimensional formulations of relativistic mechanics, then in interpreting its results we must go back to its very sources.

Let us recall briefly what we did at the beginning of Ch. 5. We *defined* the 4-momentum vector as the product of the 4-velocity and a scalar, the rest mass: $\vec{P} = m\vec{V}$. As for the equation of motion, the derivative $d\vec{P}/d\tau$ has entered its left-hand part. It can be seen from this (see also § 5.1) that the relativistic factor γ under the derivative sign in (5.37b) appeared because in 4-space-time we employ invariant proper time instead of non-invariant coordinate time. The first three components of $\vec{P}$ include simply the first three components of the 4-velocity $\vec{V}$, which bear no relation to dynamics. Thus, the factor γ refers to the properties of 4-space-time, not to the internal state of the particle.

When we introduced a scalar, the rest mass, then in 4-space the quantity obtained exact transformational properties; in other words, it is at once possible to indicate the law according to which it changes in passing from one inertial frame of reference to another. The rest mass is a scalar, i.e. invariant. This is a very important point. Since $\vec{P}^2 = -m^2c^2$ (see (5.47)), it is apparent that the rest mass is proportional to the square of the absolute value of the energy-momentum 4-vector of a particle.

As in classical mechanics, we want to associate the mass with the properties of the particle itself, in which case the only reasonable method of introducing the mass is in terms of the rest mass. One could, of course, say that a particle's acceleration to relativistic velocities causes changes in its internal properties, notably mass. But even if we do not touch the particle and introduce another inertial reference frame, the particle's equation of motion will remain (5.37b). Thus, taking "relativistic mass" at face value, it

increases for no physical reason. Such a result is hardly satisfactory.

There is no need to determine the "relativistic mass" experimentally. Only microparticles actually attain relativistic velocities, and the rest mass is sufficient to identify them. It is easily found if we determine the particle's energy and momentum from equation (5.50):

$$m^2c^2 = \frac{\mathscr{E}^2}{c^2} - p^2.$$

This is just what is done in high-energy physics.

But perhaps the "dependence of mass on velocity" can be verified directly? We should first note that no unique dependence of mass on velocity follows from the mechanics of special relativity. As pointed out in § 5.3, the essence of the matter is that, unlike Newtonian mechanics, in relativistic mechanics the directions of acceleration and force do not, in the most general case, coincide. In Newtonian mechanics a body's mass can be determined from the ratio of the magnitude of the force to the magnitude of the acceleration it imparts to the body: $m = F/(dv/dt)$. If we similarly determine the mass in relativistic mechanics we arrive at the mass tensor (regarding tensors see Appendix I, § 3). Indeed, let us rewrite (5.38) in the form ($\alpha = 1, 2, 3$)

$$m\frac{dv_\alpha}{dt} = \frac{1}{\gamma}\left[F_\alpha - \frac{v_\alpha}{c^2}(F_\beta v_\beta)\right] = \frac{1}{\gamma}\left(\delta_{\alpha\beta} - \frac{v_\alpha v_\beta}{c^2}\right)F_\beta,$$

whence we see that the acceleration is a linear vector function of the force, the factors of which (i.e. the tensor components) depend upon the velocity of the body. These factors define the inverse mass tensor:

$$m^{-1}_{\alpha\beta} = \frac{1}{m\gamma}\left(\delta_{\alpha\beta} - \frac{v_\alpha v_\beta}{c^2}\right).$$

The appearance of the mass tensor has a simple physical meaning: the magnitude of the acceleration depends upon the mutual directions of the force and the velocity. A particle's velocity represents a kind of preferred direction. For simplicity's sake we direct axis 1 along the velocity. Then $v_\alpha = \delta_{1\alpha}v$, where v is the absolute value of the velocity. We have

$$m^{-1}_{\alpha\beta} = \frac{1}{m\gamma}\left(\delta_{\alpha\beta} - \frac{v^2}{c^2}\delta_{1\alpha}\delta_{1\beta}\right),$$

or

$$m^{-1}_{\alpha\beta} = \frac{1}{m\gamma}\begin{pmatrix} \gamma^{-2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We find the mass tensor according to the conventional rule (the co-factors divided by the magnitude of the determinant):

$$m_{\alpha\beta} = m\gamma^3 \begin{vmatrix} 1 & 0 & 0 \\ 0 & \gamma^{-2} & 0 \\ 0 & 0 & \gamma^{-2} \end{vmatrix},$$

whence it is obvious that the principal values of the mass tensor $m_{\alpha\beta}$ are one "longitudinal" mass and two "transverse" masses.

But the "mass tensor" appeared solely from the desire to introduce the "dependence of mass on velocity"; the appearance of such a tensor is obviously unjustified: the rest mass is quite adequate for interpreting any results.

Thus, the answer to the question regarding the "experimental detection" of the dependence $m\gamma$ is that individual components of the tensor $m_{\alpha\beta}$ can be determined. In particular, the "transverse mass" is easily found in the motion of a charged particle in a magnetic field (see § 5.5); the "longitudinal mass" is obtained in the motion of a charged particle without an initial velocity in a uniform electric field. But if we speak of all experiments with relativistic particles in general, the simplest thing is to say that they confirm the relativistic equation of motion. In the two considered special cases when the equation of motion resembles the Newtonian (the directions of acceleration and force coincide) the equation of motion indeed looks as though the change in mass is due to the velocity. It is, however, different for the two cases. In all other cases the equation of motion is substantially different from the Newtonian. Without taking this into account one can come up against some paradoxes (see § 8.2).

If we employ only the rest mass in relativistic mechanics, it is significant from the methodological point of view that the concept of mass introduced at school remains the same. The concept of rest mass is "visualized" in the usual way from Newtonian mechanics; it is the same mass that enters the relativistic relationships of dynamics, though it can no longer be defined as the ratio of force to acceleration, but it can be defined according to (5.68). The rest mass is simply the mass used in Newtonian mechanics.

Indeed, at $\beta \ll 1$ equation (5.37b) becomes (5.37a), insofar as in this case $\gamma \approx 1$. And the determination of mass in Newtonian mechanics presents no difficulties. It is more important to stress that in relativistic mechanics the properties of space-time come into play and the laws of mechanics change, but we retain the invariance of the rest mass as a characteristic of a particle.

Sometimes attempts are made to link the increase in the energy of a particle (or a system) with an increase in mass. This is also a redundant interpretation. The dependence (5.46) shows that all

forms of energy increase equally when a particle (system) is examined not in its proper frame of reference.

The transformational properties of "relativistic mass" are also highly unsatisfactory. "Relativistic mass", which is proportional to the energy of a particle, should transform as the fourth component of the energy-momentum 4-vector. As distinct from it, the rest mass is, as mentioned before, an invariant which, like charge, characterizes an elementary particle. It is sometimes pointed out that the rest mass can change (see § 5.6), whereas the "relativistic mass" is always conserved, as long as the energy conservation law holds. But at the same time the conservation of relativistic mass yields absolutely nothing in comparison with the law of conservation of energy: it is simply a corollary of that law. Thus, the law of "conservation" of relativistic mass is a redundant equation.

The introduction of the relativistic mass of particles and the law of its "conservation" leads to the introduction of a "photon mass", $h\nu/c^2$. In § 7.6 we specifically dealt with the inexpediency of employing this quantity.

As is known, one can take as the primary principles of the special theory of relativity, not Einstein's two postulates but, for example, his first postulate and the mass-velocity dependence [32]. Formally, special relativity can be developed on such a basis. But this, of course, does not add clarity to the physical meaning of relativistic mass. It is worth emphasizing that Einstein's postulates possess a clear advantage over other possible postulates since they permit a direct physical interpretation and explicitly stress relativistic features in determining the coordinates of an event.

Sometimes it is pointed out that many eminent physicists introduced the relativistic mass. This is hardly a potent argument, the truth being that most leading physicists were against it [9, 11, 34, 35]. It is of interest that after lengthy discussions of relativistic mass Robert Feynman wrote that, strange as it may seem, the equation $m = m_0/\sqrt{1 - v^2/c^2}$ is rarely employed in practice. Instead, there are two irreplaceable relationships which are easily proved: $\mathscr{E}^2 - P^2c^2 = M_0^2c^4$ and $Pc = \mathscr{E}v/c$.

Summing up, we can say that the invariant rest mass possesses indubitable advantages, whereas relativistic mass is a source of numerous misunderstandings, while adding nothing of substance.

V. Non-inertial frames of reference. The special theory of relativity and the advance to gravitational theory (the general theory of relativity). The laws of dynamics help us single out inertial systems among all other possible frames of reference. We have defined inertial systems as those in which all three of Newton's classical laws hold. Newton's third law explicitly stresses that force is a consequence of interaction of bodies. In non-inertial

systems it is no longer possible to preserve all three laws. If the second law is retained we must introduce forces which do not satisfy the third law: inertia forces. Let us examine two examples to recall how this is done.

In the STR a reference frame is a rigid body; the most general motion of a rigid body is a combination of translational motion and rotation. Arbitrary motion of a non-inertial frame of reference relative to an inertial system is a composition of accelerated trans-

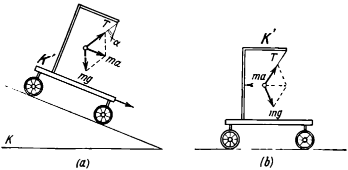


Fig. S.7. (*a*) A buggy with a suspended bead rolls down an inclined plane with constant acceleration *a*. In steady motion the thread of the pendulum is deflected somewhat from the normal to the inclined plane The force of gravity and tension of the thread combine to yield a resultant imparting to the bead the acceleration, *a*, needed for its motion together with the buggy This is the reasoning of an observer in the inertial system *K* connected with the "stationary" inclined plane. In this system the law of Newtonian dynamics holds: acceleration is caused only by forces (*b*) The same buggy and bead are examined from the point of view of system *K'* connected with the buggy, which is moving with acceleration This is a non-inertial system and the inertia force $-ma$ must be introduced. The bead is at rest relative to the buggy, hence the resultant of the three forces acting on it — gravity, the tension of the thread, and the inertia force — must vanish. It will readily be observed that the angle of deflection of the thread from the normal to the plane of the buggy is the same as in case (*a*), as it should be.

lational motion and rotation (uniform or accelerated). Uniform translational motion leaves us within the confines of inertial systems. Our examples refer to accelerated translational motion and uniform rotation.

Example 1. A buggy of mass M is rolling without friction down an inclined plane. Suspended from a bracket on the buggy is a heavy bead of mass m. Determine the angle between the thread supporting the bead and the normal to the inclined plane (Fig. S.7) for the case of steady motion.

(a) Reasoning from the point of view of the inertial coordinate system K (connected with the inclined plane). The buggy is moving with uniform acceleration $a = g \sin \alpha$ directed parallel to the

inclined plane. If the bead is at rest relative to the buggy it must be subject to the same acceleration. But acceleration is due to force, and acting on the bead are only two forces: gravity and the tension of the thread. For them to yield a resultant parallel to the inclined plane they must be at an angle. Knowing the direction of acceleration of one of them (the force of gravity) it is simple graphically to construct the direction and magnitude of the second force (the tension of the thread).

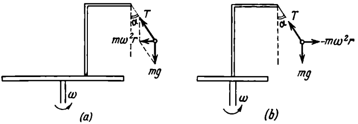


Fig. S.8. (a) A suspended bead rotates together with the turntable of a centrifugal machine. In steady motion the thread of the pendulum is deflected somewhat from the vertical away from the axis of rotation Compounded, the tension of the thread and the force of gravity yield a resultant force which gives the bead the centripetal force equal to $m\omega^2r$ and directed toward the rotation axis, needed to make the bead rotate together with the turntable. This is the reasoning of an observer in an inertial system K located outside the turntable. In this system Newton's second law holds, and the centripetal acceleration is produced by the compounding of two "real" forces (b) An observer on the turntable, that is, in the non-inertial system K', will describe the same phenomenon differently. In the K' system the bead is at rest, hence the sum of all acting forces must be zero But now there are three forces· the tension of the thread, the force of gravity, and the inertia force, equal to $-m\omega^2r$ Compounded, they yield the resultant force equal to zero. From the reasoning (b) it follows that angle α has the same value as in (a).

(b) Reasoning from the point of view of the non-inertial coordinate system K' (connected with the buggy and the bead). In this system the bead is simply at rest, hence the resultant of all the forces acting on it is zero. But in addition to the force of gravity and the tension of the thread we must take into account the inertia force $-ma$. It will be readily observed that we arrive at the same result.

Example 2. A bead of mass m suspended on a thread is placed on a centrifugal machine rotating at constant angular velocity ω (Fig. S.8). The bead is at a distance r from the rotation axis. Determine the angle of deflection of the thread from the vertical.

(a) Reasoning from the point of view of the inertial coordinate system K connected with the "stationary" stand of the centrifugal machine. For the bead to move together with the thread it must experience a centripetal acceleration $m\omega^2r$. For this the thread

must deviate from the vertical; then the resultant of the force of gravity and the tension of the thread for a given deflection from the vertical can make for the required centripetal acceleration.

(b) Reasoning from the point of view of the non-inertial coordinate system K' connected with the turntable. In this system the bead is at rest, hence the resultant of all the forces acting on it is zero. In addition to the force of gravity and the tension of the thread we must introduce the centrifugal inertia force $-m\omega^2 r$. The angle of deflection of the thread from the vertical, α, is, of course, in both cases the same.

These two examples show how inertia forces can be used to preserve Newton's second law in non-inertial coordinate systems.

These examples do not include certain other types of "inertia forces", but their essential features are apparent from them. Inertia forces are proportional to the "inert" mass of a body; they are either constant over all space (Example 1) or they increase infinitely with infinite recession from the axis of rotation (Example 2).

Galileo already knew that all bodies on Earth fall at the same rate, i.e. that the force of gravity imparts them the same acceleration. But inertia forces possess the same property. Thus, material bodies react identically to inertia forces and gravity forces. And another peculiarity of gravity forces is known from experience: there is no shielding from them (it is, in principle, possible to get rid of all other forces). That is why no direct experimental verification of Newton's first law is possible on or near the Earth. Newton himself pointed out that to verify that law one would have to get to a place where there are no gravitational fields; that is why it was stressed in Chapter 1 that Newton's first law is a postulate.

Einstein's gravitational theory possibly originated when he got the idea of the equality of all frames of reference. It seems to contradict everything discussed in this book, which has repeatedly emphasized the special role of inertial frames of reference. But let us not be in too great a hurry.

If it is impossible to get rid of either gravity or inertia, we can try to regard inertia and gravity as different aspects of the same phenomenon. Then Newton's first law must be formulated differently. The first part of Newton's statement of the law remains the same: free motion of a body is motion with no forces acting on the body, gravitation being excluded from the category of "force". Formerly, according to Newton, free motion meant uniform motion in a straight line. Now, according to Einstein, free motion is inertial and under the action of gravity forces. Gravitation is no longer a force. Now the action of forces is considered only when a body's motion is deflected from free motion, which in the Newtonian scheme was called free fall. According to Einstein, inertia and

gravity together condition "free" motion, they constitute its "background".

Of course, free motion in the Einsteinian sense is by no means in a straight line. In Euclidean geometry (on which Newtonian mechanics rests) a straight line is the shortest distance between two given points (or, as mathematicians call it, a geodesical line). We shall have to recall this a little later.

Let us get back to the conclusions obtained at the beginning of this Supplement. Passing to non-inertial reference frames simulated the appearance of inertia forces proportional to the inert mass of the body. If we recall that the gravitational mass and inert mass are equal (or proportional), it becomes apparent from the first example that passing to a reference frame in translational motion, but with acceleration, simulates the appearance of a uniform gravitational field of magnitude $-ma$. From the second example it can be seen that going over to a uniformly rotating frame of reference also leads to the appearance of a field of force proportional to the body's mass. In the general case, it can be asserted that passing to non-inertial frames of reference simulates a gravitational field. These fields have a peculiarity that distinguishes them from "genuine" gravitational fields: they do not vanish at infinity, but they do vanish in passing over to inertial frames of reference.

We have seen that a transition from one inertial frame of reference to another does not affect the square of the interval between events

$$ds^2 = c^2\,dt^2 - dx^2 - dy^2 - dz^2. \qquad \text{(S.V.1)}$$

In passing from an inertial to a non-inertial reference frame, ds^2 changes its general form. Indeed, let us consider two examples of passing from an inertial to a non-inertial frame.

Example 1. A coordinate system K' is moving in a straight line relative to K with uniform acceleration a:

$$\begin{aligned} x &= x' + at^2/2, & dx &= dx' - at'\,dt'; \\ y &= y', & dy &= dy'; \\ z &= z', & dz &= dz'; \\ t &= t', & dt &= dt'. \end{aligned}$$

If K is an inertial system and

$$ds^2 = c^2\,dt^2 - dx^2 - dy^2 - dz^2,$$

then in system K'

$$ds^2 = c^2\,dt'^2 - (dx' - at'\,dt')^2 - dy'^2 - dz'^2,$$

or

$$ds^2 = (c^2 - a^2t'^2)\,dt'^2 - 2at'\,dx'\,dt' - dx'^2 - dy'^2 - dz'^2.$$

Example 2. A uniformly rotating coordinate system (the angular velocity of rotation is Ω). From equations (A.I.10) of the Appendix I we get ($\theta = \Omega t$)

$$x = x' \cos \Omega t - y' \sin \Omega t,$$
$$y = x' \sin \Omega t + y' \cos \Omega t.$$

It is not hard to find that ds^2 transforms to the form

$$ds^2 = [c^2 - \Omega^2 (x'^2 + y'^2)]\, dt^2 + 2\Omega y'\, dx'\, dt' - 2\Omega x'\, dy'\, dt' - \\ - dx'^2 - dy'^2 - dz'^2.$$

It can be shown that in either case no time transformation can reduce ds^2 to the algebraic sum of the squares of the differentials of the four coordinates.

Thus, in the general case, going over to non-inertial systems changes the expression for the (invariant) interval between events, and in such a way that it no longer reduces to the "Galilean" form (S.V.1). Let the metric of 4-space be written down in the general case as

$$ds^2 = g_{ik}\, dx^i\, dx^k, \qquad \text{(S.V. 2)}$$

where g_{ik} depends upon all four coordinates and summation is assumed over the indices i and k. The difference between the 4-space metric (S.V.2) and the Galilean metric (S.V.1) is, according to Einstein's ideas, due to gravity. Thus, the difference between g_{ik} and the Galilean values ($g_{00} = c^2$, $g_{11} = g_{22} = g_{33} = 1$) reflects the existence of gravitational fields. But gravitational fields are associated with matter. Thus, the geometric properties (metric) of space-time are by no means invariable and they depend upon the physical objects within it. Knowledge of the space-time metric makes it possible to answer the fundamental questions that usually interest physicists. The question arises: how can these g_{ik}'s be found? Einstein was able to write a set of (non-linear) differential equations in partial derivatives which must be satisfied by ten values of g_{ik}. These values depend upon the distribution of matter and electromagnetic radiation. Einstein's equations have so far been solved only for a few special cases.

Let us briefly sum up. Passing to non-inertial reference frames causes the appearance of metric coefficients g_{ik} differing from the Galilean and simulates the appearance of a certain field of force proportional to the mass. It can therefore be assumed that the values of g_{ik} reflect the existence of a field of force similar to a gravitational field. But a similar assumption is made with respect to "true" gravitational fields, which remain in going over to inertial frames of reference. It is obvious from this that if, in an inertial frame of reference, the square of the interval is determined

according to (S.V.1), it means that there are no gravitational fields.

The difference between fields appearing in a transition to non-inertial frames of reference and true fields is that the g_{ik}'s corresponding to "true" fields cannot be reduced to the Galilean form (S.V.1) by any time or coordinate transformations. From the geometrical point of view, 4-space-time which includes gravitational fields is no longer flat. It is warped. But here we should stop and refer the reader to special literature (see, for example, [31]).

It is only necessary to note the following. In terrestrial conditions we successfully apply special relativity, i.e. employ the interval (S.V.1) along with the Newtonian theory of gravitation, i.e. we consider gravity a force. Newton's gravitational theory is explicitly non-relativistic, being a theory of action-at-a-distance. Nevertheless, the results are excellent (for example, in calculating the motions of celestial bodies). But Einstein's theory predicts that that is precisely as it should be, in certain conditions, of course. It is in "weak" gravitational fields (and within the solar system all gravitational fields are weak; there is no need to cite exact criteria) that Einstein's gravitational equations reduce to Newton's equation of gravity (Poisson's equation). As to the velocities of celestial bodies, they are always non-relativistic.

MAIN EVENTS RELATED TO THE HISTORY OF THE STR

Publication of Galileo's *Dialogue Concerning the Two Chief World Systems — Ptolemaic and Copernican*, 1632.
The first determination of the speed of light, Römer, 1676 (5?).
Publication of Newton's book *Philosophiae Naturalis Principia Mathematica*, 1687.
Discovery of the aberration of light, Bradley, 1728.
Doppler effect, 1842.
Foucault pendulum experiment, 1851.
Laboratory determination of the speed of light, Fizeau, 1849, Foucault, 1862.
Determination of the speed of light propagation in moving water, Fizeau, 1851.
Development of the theory of electromagnetic field, Maxwell, 1856-1864.
Michelson's first experiment, 1881.
Publication of E. Mach's book *Die Mechanik in ihrer Entwicklung*, 1883.
Improved Michelson's experiment, 1887.
Discovery of radioactivity, Becquerel, 1896.
Discovery of electron, J. J. Thomson, 1894-1896.
Kaufmann's study of the movement of relativistic particles in electromagnetic field, 1902.
Lorentz's papers devoted to the electrodynamics of moving bodies, 1892-1904.
Poincaré's papers on relativism, 1895-1905.
Poincaré's speech in St. Louis, 1904.
Einstein's paper "On the Electrodynamics of Moving Bodies", Ann. d. Phys. **17**, 891 (1905).
Minkowski's lecture on space and time, Phys. Zs. **10**, 104 (1909).

Main Contributors to the Development of Space and Time Science

Copernicus (1473-1543)	Doppler (1803-1853)
Galileo (1564-1642)	Maxwell (1831-1879)
Kepler (1571-1630)	Lorentz (1853-1928)
Descartes (1596-1650)	Poincaré (1854-1912)
Huygens (1629-1695)	Minkowski (1864-1909)
Newton (1643-1727)	Einstein (1879-1955)

APPENDIX I

Here we present some mathematical data needed for reading this book.

§ 1. The symmetric notation. The summation rules. When a rectilinear orthogonal (Cartesian) system of coordinates is introduced in the three-dimensional space, the unit vectors oriented along the x, y, z axes are denoted by $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ respectively. The position of any point in space is determined by the radius vector $\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$ whose components are the coordinates of the point. All directions in space being equivalent, it is expedient to introduce the symmetric notation and to write, for example, x_1, x_2, x_3 instead of x, y, z, and $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, $\boldsymbol{m}_3$ instead of $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$. Then the radius vector of a point will be written in the following form:

$$\boldsymbol{r} \equiv x_1\boldsymbol{m}_1 + x_2\boldsymbol{m}_2 + x_3\boldsymbol{m}_3 = \sum_{\alpha=1}^{3} x_\alpha \boldsymbol{m}_\alpha, \tag{A.I.1}$$

where $\sum$ denotes the summation carried out in this case over α running from 1 to 3. The summation sign can be omitted, having once and for all stipulated that the two identical Greek letter indices appearing in one side of an equation imply the summation from 1 to 3.

The special theory of relativity employs a four-dimensional space with four coordinates $x_1 = x$, $x_2 = y$, $x_3 = z$, $x_4 = ict$. The radius vector and other vectors have in this case four components. The summation rule is valid, but the summation is carried out for Latin letter indices running over all values from 1 to 4. For example,

$$A_iB_i \equiv \sum_{i=1}^{4} A_iB_i = A_1B_1 + A_2B_2 + A_3B_3 + A_4B_4.$$

So, the concise summation rule prescribes the summation over two identical indices appearing in one side of an equation; in the case of Latin letter indices the summation runs over the values from 1 to 4.

Let us go back to the three-dimensional case. The radius vector (A. I. 1) can be written in an abbreviated form as $\boldsymbol{r} = x_\alpha \boldsymbol{m}_\alpha$ and arbitrary vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ as

$$\boldsymbol{a} = a_\alpha \boldsymbol{m}_\alpha = a_\beta \boldsymbol{m}_\beta = a_\gamma \boldsymbol{m}_\gamma, \quad \boldsymbol{b} = b_\alpha \boldsymbol{m}_\alpha = b_\beta \boldsymbol{m}_\beta = b_\gamma \boldsymbol{m}_\gamma.$$

The same equality is written out several times in order to show that the summation indices are "mute", i.e. the summation can be carried out over any letter index with the result remaining constant.

To illustrate the application of the abbreviated notation let us derive the formula for the scalar product of two vectors a and b. On the one hand

$$\boldsymbol{ab} = a_\alpha \boldsymbol{m}_\alpha b_\beta \boldsymbol{m}_\beta = a_\alpha b_\beta \boldsymbol{m}_\alpha \boldsymbol{m}_\beta. \tag{A.I.2}$$

Here we take into account that the summation rule pertains to two identical indices; there are two summations to be carried out in Eq. (A.I.2), each of which being performed over its own letter index. On the other hand, unit vectors are mutually orthogonal; therefore, each vector yields unity when multiplied by itself and zero when multiplied by any other vector. Consequently,

$$\boldsymbol{m}_\alpha \boldsymbol{m}_\beta = \begin{cases} 1, & \alpha = \beta, \\ 0, & \alpha \neq \beta. \end{cases} \tag{A.I.3}$$

It is convenient to introduce the Kronecker delta possessing exactly these properties:

$$\delta_{\alpha\beta} = \begin{cases} 1, & \alpha = \beta, \\ 0, & \alpha \neq \beta. \end{cases} \tag{A.I.4}$$

This symbol differs from zero only at $\alpha = \beta$, and any summation involving this symbol yields the simple result: $a_\alpha \delta_{\alpha\beta} = a_\beta$. Indeed

$$a_\alpha \delta_{1\alpha} = a_1 \delta_{11} + a_2 \delta_{12} + a_3 \delta_{13} = a_1.$$

Now we can easily complete the derivation of Eq. (A. I. 2)

$$\boldsymbol{ab} = a_\alpha b_\beta \boldsymbol{m}_\alpha \boldsymbol{m}_\beta = a_\alpha b_\beta \delta_{\alpha\beta} = a_\alpha b_\alpha$$

to obtain the conventional formula for a scalar product of vectors.

The identical indices over which the summation is carried out may appear in a numerator and denominator of a fraction. The summation rule is valid in this case as well. Let us write, for example, the expression for a gradient of the function f and a divergence of the vector $\boldsymbol{a}$:

$$\operatorname{grad} f = \frac{\partial f}{\partial x_\alpha} \boldsymbol{m}_\alpha = \frac{\partial f}{\partial x_1} \boldsymbol{m}_1 + \frac{\partial f}{\partial x_2} \boldsymbol{m}_2 + \frac{\partial f}{\partial x_3} \boldsymbol{m}_3 =$$

$$= \frac{\partial f}{\partial x} \boldsymbol{i} + \frac{\partial f}{\partial y} \boldsymbol{j} + \frac{\partial f}{\partial z} \boldsymbol{k}, \quad \operatorname{div} \boldsymbol{a} = \frac{\partial a_\alpha}{\partial x_\alpha} = \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}.$$

When a Greek (or Latin) letter index appears alone, it is meant to be "free", taking on any value out of the three (or four) possible ones. For example, b_α denotes one of the coordinates of the vector $\boldsymbol{b}$, that is b_1, b_2 or b_3.

§ 2. **The transformation of coordinates in the case of a rotation of the Cartesian system of coordinates.** Let the radius vector of the point M be expressed as $\boldsymbol{r} = x_\alpha \boldsymbol{m}_\alpha$ in the "old" coordinate system. After the rotation of the coordinate system the radius vector of the same point M will be written in the "new" coordinate system as $\boldsymbol{r} = x'_\beta \boldsymbol{m}'_\beta$ where x'_β are the coordinates of the point in the system after the rotation and $\boldsymbol{m}'_\beta$ are the new unit vectors. It is not difficult to define the relationship between the coordinates of the old and new systems. Let us write down the equality expressing the "conservation" of the vector $\boldsymbol{r}$,

$$x_\alpha \boldsymbol{m}_\alpha = x'_\beta \boldsymbol{m}'_\beta,$$

and multiply its both sides by $\boldsymbol{m}'_\gamma$, i.e. an arbitrary unit vector of the new coordinate system. The left-hand side yields

$$x_\alpha \boldsymbol{m}_\alpha \boldsymbol{m}'_\gamma = x_\alpha a_{\alpha\gamma};$$

here we used the designation $\boldsymbol{m}_\alpha \boldsymbol{m}'_\gamma = \cos(\widehat{\boldsymbol{m}_\alpha, \boldsymbol{m}'_\gamma}) = a_{\alpha\gamma}$; thus, $a_{\alpha\gamma}$ represents a cosine of the angle between the vector $\boldsymbol{m}_\alpha$ of the old system and the vector $\boldsymbol{m}'_\gamma$ of the new one. On the right-hand side we obtain the following chain of equalities:

$$x'_\beta \boldsymbol{m}'_\beta \boldsymbol{m}'_\gamma = x'_\beta \delta_{\beta\gamma} = x'_\gamma.$$

Thus,

$$x'_\gamma = a_{\alpha\gamma} x_\alpha \quad (\gamma = 1,\ 2,\ 3). \tag{A.I.5}$$

The new coordinates are expressed via the old ones linearly, the coefficients being the cosines of the angles between the old and the new coordinate axes. We have to find the coefficients of the expansion of the old unit vectors in terms of the new ones. Let us expand the old vector $\boldsymbol{m}_\gamma$ via the new ones:

$$\boldsymbol{m}_\alpha = a^{\bullet}_{\alpha\mu} \boldsymbol{m}'_\mu, \tag{A.I.6}$$

where $a^{\bullet}_{\alpha\mu}$ are unknown coefficients. To find them, let us multiply both sides of this equation by $\boldsymbol{m}'_\gamma$. Similarly to the foregoing formula

$$a_{\alpha\gamma} = a^{\bullet}_{\alpha\mu} \delta_{\mu\gamma} = a^{\bullet}_{\alpha\gamma}. \tag{A.I.7}$$

We have obtained an obvious result: the coefficients of the expansion of the unit vector $\boldsymbol{m}_\alpha$ with respect to the new unit vectors are the cosines $a_{\alpha\gamma}$.

The cosines of the angles between the old and new vectors can be combined in the matrix:

$$a_{\alpha\beta} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}; \qquad \text{(A.I.8)}$$

in the designation $a_{\alpha\beta}$ the first index, α, indicates the row, and the second, β, the column of the matrix (A. I. 8). Thus, the transformation of coordinates is determined by nine coefficients $a_{\alpha\beta}$. It is known, however, that the position of any solid body (in our case the coordinate system) whose one point is motionless can be defined by three parameters (three Eulerian angles). Whence it is clear that there are only three independent coefficients among the nine coefficients of the matrix $a_{\alpha\beta}$. It is easy to find the requisite relations between the coefficients $a_{\alpha\beta}$. Indeed, the rotation of the coordinate system does not affect the distance from the origin to any point: $r^2 = x_\beta^2 = x_\alpha'^2$. But $x'_\alpha = a_{\beta\alpha} x'_\beta$. To square this expression, we have to multiply the sums whose summation indices should be different:

$$x_\alpha'^2 = x'_\alpha x'_\alpha = a_{\beta\alpha} x_\beta a_{\gamma\alpha} x_\gamma = a_{\beta\alpha} a_{\gamma\alpha} x_\beta x_\gamma.$$

But, on the other hand, this expression is equal to x_β^2. This can be only if

$$a_{\beta\alpha} a_{\gamma\alpha} = \delta_{\beta\gamma} \quad (\beta,\ \gamma = 1,\ 2,\ 3). \qquad \text{(A.I.9)}$$

Although at first glance there are nine conditions here, these equations do not change when their indices β and γ are interchanged. Consequently, there are only six independent equations here. Each of them represents a product of the βth and γth rows of the matrix (A.I.8). (The matrix row multiplication consists in the summation of pairwise products of the respective elements.) Eq. (A.I.9) means that the product of any row by itself is equal to unity and by any other row to zero. Since the order of the row multiplication makes no difference, e.g. the product of the first row by the second row is equal to the product of the second row by the first one, the number of independent equations is equal to six and not to nine, as it was indicated.

The relations obtained are best illustrated by the example of the rotation of axes in the coordinate plane $(x_1,\ x_2)$. In this case

$$x'_1 = a_{11} x_1 + a_{21} x_2, \quad x'_2 = a_{12} x_1 + a_{22} x_2.$$

In accordance with Fig. (A.1),

$$a_{11} = \cos\theta, \qquad a_{21} = \sin\theta,$$
$$a_{12} = -\sin\theta, \quad a_{22} = \cos\theta,$$

and therefore,

$$x_1' = x_1 \cos\theta + x_2 \sin\theta, \quad x_2' = -x_1 \sin\theta + x_2 \cos\theta. \qquad \text{(A.I.10)}$$

Passing to the conventional notation $x_1 = x$, $x_2 = y$, we obtain the well-known equations of analytical geometry:

$$x' = x\cos\theta + y\sin\theta, \quad y' = -x\sin\theta + y\cos\theta.$$

We used these equations in the derivation of the Lorentz transformation. We have obtained the equations of the direct transition (from the unprimed to primed system).

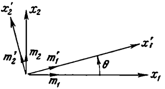


Fig. A.1. The illustration of the general formulae for the coordinate transformation by means of a rotation of the Cartesian system in a plane. Such a rotation is defined by one parameter θ. The angles between the old and new coordinate vectors are seen in the figure.

The reverse transition equations are obtained in much the same fashion. We shall write them out together with the direct transition equations:

$$\begin{aligned} x_\alpha' &= a_{\beta\alpha} x_\beta', \quad m_\alpha' = a_{\beta\alpha} m_\beta, \\ x_\alpha &= a_{\alpha\beta} x_\beta', \quad m_\alpha = a_{\alpha\beta} m_\beta', \end{aligned} \qquad \text{(A.I.11)}$$

at the same time

$$a_{\alpha\beta} = m_\alpha m_\beta' = \cos(\widehat{m_\alpha m_\beta'}).$$

Surely, the reverse transition equations (see (A.I.11)) can be obtained automatically from the direct transition equations by exchanging primed and unprimed quantities and replacing the angle θ by $-\theta$ (which corresponds to the rotation in the opposite direction).

How are the vector components transformed on the coordinate transformation? This can be easily found by the same technique that we used when deriving equations for the transformation of coordinates. We may not do this, however, having noticed that the coordinates are also the components of a vector, that is, of the radius vector. Therefore, it is clear that the vector components are transformed as the coordinates are, i.e.

$$b_\alpha = a_{\alpha\beta} b_\beta', \quad b_\alpha' = a_{\beta\alpha} b_\beta. \qquad \text{(A.I.12)}$$

As we have already mentioned, the four-dimensional (pseudo-Euclidean) space used in the special theory of relativity includes formally one imaginary coordinate associated with time

$$x_1 = x, \quad x_2 = y, \quad x_3 = z, \quad x = ict.$$

The Lorentz transformation corresponds to linear transformations in this space:

$$x_i' = \alpha_{ki} x_k, \quad x_i = \alpha_{ik} x_k', \tag{A.I.13}$$

and

$$\alpha_{ik} = \begin{pmatrix} \Gamma & 0 & 0 & -i\mathrm{B}\Gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ i\mathrm{B}\Gamma & 0 & 0 & \Gamma \end{pmatrix}, \quad \Gamma = \frac{1}{\sqrt{1 - \frac{V^2}{c^2}}}, \tag{A.I.14}$$

where V is the relative velocity of two reference frames.

The Lorentz transformation coefficients α_{ik} satisfy the following conditions:

$$\alpha_{kl} \alpha_{km} = \delta_{lm}. \tag{A.I.15}$$

These equalities mean that the product of the Lorentz transformation matrix rows yields unity when a row is multiplied by itself, and zero when a row is multiplied by any other row.

Let us calculate the determinant of the Lorentz matrix *

$$\begin{vmatrix} \Gamma & 0 & 0 & -i\mathrm{B}\Gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ i\mathrm{B}\Gamma & 0 & 0 & \Gamma \end{vmatrix} = \Gamma^2 + (-i\mathrm{B}\Gamma)(-i\mathrm{B}\Gamma) = \Gamma^2 (1 - \mathrm{B}^2) = 1$$

(the simplest way is to expand the determinant into the first row elements). The Lorentz matrix determinant proves to be equal to unity. This means that we deal with the proper Lorentz transformation, i.e. we stay in the systems of clockwise triads of unit coordinate vectors and do not pass to the systems of anticlockwise triads.

Let us write the direct and the inverse Lorentz transformation for coordinates in full:

$$x_1 = \Gamma (x_1' - i\mathrm{B}x_4'), \quad x_2 = x_2', \quad x_3 = x_3', \quad x_4 = \Gamma (x_4' + i\mathrm{B}x_1'), \tag{A.I.16}$$

$$x_1' = \Gamma (x_1 + i\mathrm{B}x_4), \quad x_2' = x_2, \quad x_3' = x_3, \quad x_4' = \Gamma (x_4 - i\mathrm{B}x_1). \tag{A.I.17}$$

* Some data on determinants can be found in § 6 of this Appendix. This determinant is calculated there as well.

As to 4-vector components, they are transformed as coordinates, and, consequently, we get for the vector $\vec{A}$ (A_1, A_2, A_3, A_4) (the arrow over a letter denotes a 4-vector):

$$A'_i = \alpha_{ki} A_k, \qquad A_i = \alpha_{ik} A_k, \tag{A.I.18}$$

$$A'_1 = \Gamma (A_1 + iBA_4), \quad A'_2 = A_2, \quad A'_3 = A_3, \quad A'_4 = \Gamma (A_4 - iBA_1), \tag{A.I.19}$$

$$A_1 = \Gamma (A'_1 + iBA'_4), \quad A_2 = A'_2, \quad A_3 = A'_3, \quad A_4 = \Gamma (A'_4 + iBA'_1). \tag{A.I.20}$$

From Eq. (A.I.15) it follows that the scalar product of two 4-vectors is invariant under the Lorentz transformation. Indeed, if, $A_i = \alpha_{ik} A'_k$, $B_i = \alpha_{im} B'_m$ then

$$\vec{A}\vec{B} = A_i B_i = \alpha_{ik} A'_k \alpha_{im} B'_m = \alpha_{ik} \alpha_{im} A'_k B'_m = \delta_{km} A'_k B'_m = A'_k B'_k.$$

The comparison of the second and the last link of the written chain of equalities proves the invariance of the scalar product $\vec{A}\vec{B}$.

§ 3. The tensors. Vector quantities are a particular case of tensors, mathematical quantities of much more complicated nature. Prior to passing over to them, let us point out the most essential points in the definition of a vector. In a given coordinate system a vector represents a directed line segment characterized by its own coordinates. But since a coordinate system is chosen at will, the vector coordinates are random. What is important, however, is that using the vector coordinates determined in one Cartesian system, we can find its Cartesian coordinates in any other system by means of Eq. (A.I.5). These are the transformation equations which define the vector. Thus, the vectorial nature of quantities is revealed under the transformation of coordinates.

To clear up the concept of a tensor by means of a specific example, we shall recall how the relationship between the electric induction vector $\boldsymbol{D}$ and the external electric field strength $\boldsymbol{E}$ is introduced in electrostatics. By and large, the relation $\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{E}) = \boldsymbol{D}(E_1, E_2, E_3)$ is unknown. Let us expand $\boldsymbol{D}$ into the components D_α:

$$\boldsymbol{D}(\boldsymbol{E}) = D_\alpha (\boldsymbol{E}) \boldsymbol{m}_\alpha = D_\alpha (E_1, E_2, E_3) \boldsymbol{m}_\alpha.$$

Assuming that in the absence of the external field ($\boldsymbol{E} = 0$) the vector $\boldsymbol{D}$ is also equal to zero ($\boldsymbol{D}(0) = 0$) and assuming the external field to be small as compared to electric forces acting between molecules of a substance, we can expand the unknown vector

function $\boldsymbol{D}$ in a Taylor series:

$$D_1 = \frac{\partial D_1(0)}{\partial E_1} E_1 + \frac{\partial D_1(0)}{\partial E_2} E_2 + \frac{\partial D_1(0)}{\partial E_3} E_3 + \ldots = \frac{\partial D_1(0)}{\partial E_\beta} E_\beta + \ldots, \tag{A.I.21}$$

$$D_2 = \frac{\partial D_2(0)}{\partial E_\beta} E_\beta + \ldots, \quad D_3 = \frac{\partial D_3(0)}{\partial E_\beta} E_\beta + \ldots,$$

where the summation is carried out over the index β. Due to the smallness of the field $\boldsymbol{E}$ its components E_1, E_2 and E_3 are also small (in fact, this is a good approximation in the case of real fields, except for the fields generated in laser beams), and we can take only the linear terms, neglecting all the others. Let us denote the constant quantities, which are the derivatives at the zero point, as follows:

$$\frac{\partial D_\alpha 0}{\partial E_\beta} = \varepsilon_{\alpha\beta}.$$

Then the expressions obtained can be written in the form

$$\begin{aligned} D_1 &= \varepsilon_{11}E_1 + \varepsilon_{12}E_2 + \varepsilon_{13}E_3, \\ D_2 &= \varepsilon_{21}E_1 + \varepsilon_{22}E_2 + \varepsilon_{23}E_3, \\ D_3 &= \varepsilon_{31}E_1 + \varepsilon_{32}E_2 + \varepsilon_{33}E_3 \end{aligned} \tag{A.I.22}$$

or in the abbreviated form

$$D_\alpha = \varepsilon_{\alpha\beta} E_\beta. \tag{A.I.23}$$

Using its components, we can easily compose the vector $\boldsymbol{D}$:

$$\boldsymbol{D} = \varepsilon_{\alpha\beta} E_\beta \boldsymbol{m}_\alpha. \tag{A.I.24}$$

The relationship between two vectors expressed by Eqs. (A.I.23) and (A.I.24) is referred to as a linear vector function; in other words, the vector $\boldsymbol{D}$ is a linear vector function of E.

Using Eq. (A.I.24), we can construct the vector $\boldsymbol{D}$ from the given vector $\boldsymbol{E}$ at each point of a dielectric in the coordinate system where the coefficients $\varepsilon_{\alpha\beta}$ are known. The coordinate system, however, is chosen randomly. The rotation of the Cartesian coordinate system changes the vector components without varying the vectors themselves. The question is how the coefficients $\varepsilon_{\alpha\beta}$ must change to maintain the relationship $\boldsymbol{D}' = \varepsilon'_{\mu\lambda} E'_\lambda \boldsymbol{m}'_\mu$ in the new system, with $\boldsymbol{D} = \boldsymbol{D}'$. This means that there must be two different expansions of the same vector:

$$\boldsymbol{D} = \varepsilon_{\alpha\beta} E_\beta \boldsymbol{m}_\alpha = \varepsilon'_{\mu\lambda} E'_\lambda \boldsymbol{m}'_\mu. \tag{A.I.25}$$

The vector components and unit vectors are known to be transformed according to Eq. (A.I.11): $E_\beta = a_{\beta\lambda} E'_\gamma$, $\boldsymbol{m}_\alpha = a_{\alpha\mu} \boldsymbol{m}'_\mu$. In

accordance with these equations the left-hand side of Eq. (A.I.25) can be rewritten in the following form (the right-hand side is left unchanged):

$$\varepsilon_{\alpha\beta} a_{\beta\lambda} a_{\alpha\mu} E'_{\lambda} m'_{\mu} = \varepsilon'_{\mu\lambda} E'_{\lambda} m'_{\mu}.$$

Comparing the coefficients of $E'_{\lambda} m'_{\mu}$ in the left-hand and right-hand sides, we obtain the transformation law for the coefficients $\varepsilon_{\alpha\beta}$:

$$\varepsilon'_{\mu\lambda} = a_{\beta\lambda} a_{\alpha\mu} \varepsilon_{\alpha\beta}. \tag{A.I. 26}$$

The comparison of this equation with the coordinate transformation law

$$x'_{\lambda} = a_{\beta\lambda} x_{\beta} \tag{A.I. 27}$$

shows that each index of $\varepsilon_{\alpha\beta}$ is transformed according to the law corresponding to the coordinate transformation law. Eq. (A.I.26) represents the tensor transformation law. The inverse transformation obviously takes the form $\varepsilon_{\mu\lambda} = a_{\lambda\beta} a_{\mu\alpha} \varepsilon'_{\alpha\beta}$.

Here is the general definition of a tensor: if in a given Cartesian coordinate system we have nine quantities $\varepsilon_{\alpha\beta}$ which, under the coordinate transformation $x^{a} = a_{\beta a} x_{a}$, are transformed according to the formulae

$$\varepsilon'_{\alpha\beta} = a_{\gamma\alpha} a_{\mu\beta} \varepsilon'_{\gamma\mu}, \tag{A.I. 28}$$

these quantities form a tensor of the second rank. It is not difficult to realize that vectors are transformed as tensors of the first rank. The rank of a tensor (or, as it is sometimes called, the valency of a tensor) is defined by the number of its indices. There are two such indices in our case. Tensors of a higher rank are almost never used in this book. A tensor is defined for the space possessing a definite number of dimensions since its transformation law involves the components of the transformation matrix. We have been considering the three-dimensional space, and the Greek letter indices α, β, γ, μ were varying from one to three.

We would like to point out two distinctive characteristics of a tensor transformation:

(1) The transformation law for the coefficients of a linear vector function (tensor) is obtained as a condition for the invariant physical relationship between vectors.

(2) Any component of a tensor in the "new" coordinate system is a linear combination of all components of this tensor in the "old" system.

As a useful special case, let us note the transformation of a three-dimensional tensor of the second rank $T_{\alpha\beta}$ whose only nonzero component T_{11} corresponds to a rotation in a plane. In the primed system the following components will differ from zero (see Eqs. (A.I.28) and (A.I.10)):

$$\begin{aligned} T'_{11} &= a_{\lambda 1} a_{\mu 1} T_{\lambda\mu} = a_{11}^2 T_{11} = T_{11}\cos^2\theta, \\ T'_{12} &= a_{\lambda 2} a_{\mu 2} T_{\lambda\mu} = a_{11} a_{12} T_{11} = -T_{11}\sin\theta\cos\theta. \end{aligned} \qquad \text{(A.I. 29)}$$

We shall come across these equations many times in the future.

The special theory of relativity deals with the four-dimensional (pseudo-Euclidean) space. We have already considered the transformation laws for 4-vectors in this space. (According to our definition, a vector is a tensor of the first rank.) The transformation rules for tensors in the 4-space remain actually the same, but the number of tensor components increases to sixteen while the summation is carried out from 1 to 4:

$$A_{ik} = \alpha_{im}\alpha_{kl}A'_{ml}, \quad A'_{ik} = \alpha_{mi}\alpha_{lk}A_{ml}. \qquad \text{(A.I. 30)}$$

A tensor is referred to as symmetric if its components satisfy the relation $A_{ik} = A_{ki}$. Such a tensor has only ten independent components. The energy-momentum-tension tensor of an electromagnetic field may serve as an example of a symmetric tensor.

A tensor is referred to as antisymmetric if its components satisfy the relation $A_{ik} = -A_{ki}$. It is clear that the elements of this tensor having identical indices ($i = k$) are equal to zero since the only quantity which is equal to itself when taken with the opposite sign is zero. Thus, an antisymmetric tensor has only six independent components (in this connection it is sometimes called a six-vector tensor). The electromagnetic field tensor provides an example of an antisymmetric tensor.

A tensor is referred to as unitary if $A_{ik} = \delta_{ik}$. It is easy to see that the tensor δ_{ik} retains its form under the Lorentz transformation in all reference frames. Indeed, if $A_{ml} = \delta_{ml}$, then according to Eq. (A.I.30)

$$A'_{ik} = \alpha_{im}\alpha_{kl}A_{ml} = \alpha_{im}\alpha_{kl}\delta_{ml} = \alpha_{im}\alpha_{km} = \delta_{ik};$$

the last transition is performed in accordance with Eq. (A.I.5).

For reference we shall write the transformation formulae for the tensor of the second rank T_{ik} in the case of the Lorentz transformation, i.e. the formulae describing the transition from the frame K' to the frame K (the inverse transition formulae are obtained by changing the sign of V and by replacing the primed quantities

by unprimed ones and vice versa):

$$\left.\begin{aligned}
T_{11} &= \alpha_{11}^2 T'_{11} + \alpha_{14}\alpha_{11}(T'_{41} + T'_{14}) + \alpha_{14}^2 T'_{44} = \\
&= \Gamma^2\{T'_{11} - i\mathrm{B}(T'_{14} + T'_{41}) - \mathrm{B}^2 T'_{44}\}, \\
T_{12} &= \alpha_{11}T'_{12} + \alpha_{14}T'_{42} = \Gamma(T'_{12} - i\mathrm{B}T'_{42}), \\
T_{13} &= \alpha_{11}T'_{13} + \alpha_{14}T'_{43} = \Gamma(T'_{13} - i\mathrm{B}T'_{43}), \\
T_{14} &= \alpha_{11}\alpha_{41}T'_{11} + \alpha_{14}\alpha_{41}T'_{41} + \alpha_{11}\alpha_{44}T'_{14} + \alpha_{11}\alpha_{44}T'_{44} = \\
&= \Gamma^2\{T'_{14} + i\mathrm{B}T'_{11} + \mathrm{B}^2 T'_{41} - i\mathrm{B}T'_{44}\}, \\
T_{21} &= \alpha_{11}T'_{21} + \alpha_{14}T'_{24} = \Gamma(T'_{21} - i\mathrm{B}T'_{24}), \\
T_{22} &= T'_{22}, \quad T_{23} = T'_{23}, \\
T_{24} &= \alpha_{41}T'_{21} + \alpha_{44}T'_{24} = \Gamma(T'_{24} + i\mathrm{B}T'_{21}), \\
T_{31} &= \alpha_{11}T'_{31} + \alpha_{14}T'_{34} = \Gamma(T'_{31} - i\mathrm{B}T'_{34}), \\
T_{32} &= T'_{32}, \quad T_{33} = T'_{33}, \\
T_{34} &= \alpha_{41}T'_{31} + \alpha_{44}T'_{34} = \Gamma(T'_{34} + i\mathrm{B}T'_{31}), \\
T_{41} &= \alpha_{11}\alpha_{41}T'_{11} + \alpha_{14}\alpha_{41}T'_{14} + \alpha_{11}\alpha_{44}T'_{41} + \alpha_{14}\alpha_{44}T'_{44} = \\
&= \Gamma^2(T'_{41} + i\mathrm{B}T'_{11} + \mathrm{B}^2 T'_{14} - i\mathrm{B}T'_{44}), \\
T_{42} &= \alpha_{42}T'_{12} + \alpha_{44}T'_{42} = \Gamma(T'_{42} + i\mathrm{B}T'_{12}), \\
T_{43} &= \alpha_{41}T'_{13} + \alpha_{44}T'_{43} = \Gamma(T'_{43} + i\mathrm{B}T'_{13}), \\
T_{44} &= \alpha_{41}^2 T'_{11} + \alpha_{41}\alpha_{44}T'_{14} + \alpha_{44}^2 T'_{44} = \\
&= \Gamma^2\{T'_{44} + i\mathrm{B}(T'_{14} + T'_{41}) - \mathrm{B}^2 T'_{11}\}.
\end{aligned}\right\} \quad (\text{A.I. }31)$$

The tensor quantities appear more often than it may seem at first glance. We shall give some examples of tensors of the second rank in the 4-space. The products of the components of two vectors $\vec{c}(c_i)$ and $\vec{b}(b_k)$ form a tensor. Indeed, let us compose the expression $A_{ik} = c_i b_k$. The transformation formulae for the vector components are known:

$$c_i = \alpha_{im} c'_m, \quad b_k = \alpha_{kl} b'_l. \quad (\text{A.I. }32)$$

Consequently,

$$A_{ik} = c_i b_k = \alpha_{im}\alpha_{kl} c'_m b'_l = \alpha_{im}\alpha_{kl} A'_{ml},$$

coinciding with the tensor transformation law (A.I.30).

Let us demonstrate that the derivative of a vector component with respect to a coordinate is transformed as a tensor component. Let us consider the vector $\vec{b}(b_i)$ and its component derivatives. It will be easier for us to write out two formulae: $b_l = a_{lm}b'_m$, $x'_l = a_{kl}x_k$, whence $\frac{\partial x'_l}{\partial x_k} = a_{kl}$ to be used in the following chain of equalities:

$$\frac{\partial b_l}{\partial x_k} = \frac{\partial b_l}{\partial x'_l}\frac{\partial x'_l}{\partial x_k} = \frac{\partial}{\partial x'_l}(a_{lm}b'_m)\frac{\partial x'_l}{\partial x_k} = a_{lm}a_{kl}\frac{\partial b'_m}{\partial x'_l}. \quad \text{(A.I. 33)}$$

The first and the last links of Eq. (A.I.33) show that the derivative $\partial b_l/\partial x_k$ is transformed according to the tensor component transformation law.

From Eq. (A.I.33) it is also seen that the vector derivative can be transformed in succession. First, we can pass from b'_m to b_l according to the formulae $b_l = a_{lm}b'_m$. Then from differentiating with respect to x'_l we can proceed to differentiating with respect to x_k. Of course, the tensor character of the transformation is retained in this process although in a rather concealed form That is exactly how they sometimes do when trying to avoid tensors in transformations of an electromagnetic field

§ 4. **The invariance of a 4-divergence and d'Alembert's operator.** Let us demonstrate the invariance of a four-dimensional divergence and d'Alembert's operator under the Lorentz transformation. We shall write out the necessary formulae in a convenient form:

$$x'_l = a_{kl}x_k, \quad \frac{\partial x'_l}{\partial x_k} = a_{kl},$$

and for the components of the vector $\vec{b}$:

$$b'_l = a_{kl}b_k, \quad b_k = a_{kl}b'_l.$$

First, let us prove that a 4-gradient is transformed as a vector. We shall consider the function $\varphi = \varphi(x'_1, x'_2, x'_3, x'_4)$ or, briefly, $\varphi = \varphi(x'_l)$. Let the coordinate transformation be defined by the formulae $x'_l = x'_l(x_1, x_2, x_3, x_4)$. Then according to the differentiation rule for composite functions

$$\frac{\partial \varphi}{\partial x_k} = \frac{\partial \varphi}{\partial x'_l}\frac{\partial x'_l}{\partial x_k} = a_{kl}\frac{\partial \varphi}{\partial x'_l}.$$

In this way we obtain the vector component transformation law (A.I.32) again.

Now let us demonstrate the invariance of the 4-divergence. The following chain of equalities proves the fact:

$$\operatorname{div}\vec{A}=\frac{\partial A_k}{\partial x_k}=\frac{\partial A_k}{\partial x'_l}\frac{\partial x'_l}{\partial x_k}=\alpha_{kl}\frac{\partial}{\partial x'_l}(A'_i\alpha_{ki})=\alpha_{ki}\alpha_{kl}\frac{\partial A'_i}{\partial x'_l}=$$

$$=\delta_{il}\frac{\partial A'_i}{\partial x'_l}=\frac{\partial A'_l}{\partial x'_l}.$$

Here we have taken into account that in accordance with Eq. (A.I.15) $\alpha_{ki}\alpha_{kl}=\delta_{il}$.

D'Alembert's operator applied to the function Φ,

$$\Box\,\Phi\equiv\Delta\Phi-\frac{1}{c^2}\frac{\partial^2\Phi}{\partial t^2}=\frac{\partial^2\Phi}{\partial x^2}+\frac{\partial^2\Phi}{\partial y^2}+\frac{\partial^2\Phi}{\partial z^2}-\frac{1}{c^2}\frac{\partial^2\Phi}{\partial t^2},$$

will be written in the form

$$\frac{\partial^2\Phi}{\partial x_l^2}=\frac{\partial^2\Phi}{\partial x_1^2}+\frac{\partial^2\Phi}{\partial x_2^2}+\frac{\partial^2\Phi}{\partial x_3^2}+\frac{\partial^2\Phi}{\partial (ict)^2}.$$

The expression $\partial^2\Phi/\partial x_l^2$, however, is the divergence of a gradient. Indeed, let $A_l=\partial\Phi/\partial x_l$, then

$$\operatorname{div}\vec{A}=\frac{\partial}{\partial x_l}\left(\frac{\partial\Phi}{\partial x_l}\right)=\frac{\partial^2\Phi}{\partial x_l^2}.$$

But the divergence of a 4-vector is an invariant of the Lorentz transformation, and consequently

$$\frac{\partial^2\Phi}{\partial x_l^2}=\frac{\partial^2\Phi}{\partial x_l'^2}. \tag{A.I. 34}$$

Any component of a 4-vector $\vec{\Phi}(\Phi_k)$ can be taken in the capacity of the function Φ. Let us assume the respective components of the two 4-vectors $\vec{\Phi}(\Phi_k)$ and $\vec{s}(s_k)$ to be related in the frame K by the following equation:

$$\Box\,\Phi_k\equiv\frac{\partial^2\Phi_k}{\partial x_l^2}=-\mu_0 s_k. \tag{A.I. 35}$$

Then it follows from Eq. (A.I.34) that the relation

$$\Box'\Phi_k=\frac{\partial^2\Phi_k}{\partial x_l'^2}=-\mu_0 s_k \tag{A.I. 36}$$

is valid provided Eq. (A.I.35) is. Multiplying the left-hand and right-hand sides of Eq. (A.I.36) by the constant factor α_{km} and summing over k, we immediately get

$$\Box'\Phi'_m=-\mu_0 s'_m, \tag{A.I. 37}$$

since $a_{km}\Phi_k = \Phi'_m$. The intercomparison of Eqs. (A.I.35) and (A.I.37) shows that in the frame K' we have exactly the same equation as in K in which the primed and unprimed quantities are interchanged.

§ 5. **The convolution ("rejuvenation") of tensor indices.** In the tensor calculus there is an operation causing the rank of a tensor to become lower. In the case of a tensor of the second rank this operation consists in summing up the tensor components possessing two identical indices. It is noteworthy that such an operation gives rise to the invariant expression. In the case of tensors of a higher rank the convolution results in the tensor's rank getting lower by two.

We can demonstrate this property very easily. Let us write the tensor component transformation formula

$$A_{ik} = a_{im}a_{kl}A'_{ml}$$

and sum up the components A_{ik} possessing the identical indices having put $i = k$. Then due to Eq. (A.I.15)

$$A_{ii} = A_{11} + A_{22} + A_{33} + A_{44} = a_{im}a_{il}A'_{ml} = \delta_{ml}A'_{ml} = A'_{mm}.$$

Thus,

$$A_{11} + A_{22} + A_{33} + A_{44} = A'_{11} + A'_{22} + A'_{33} + A'_{44}. \quad \text{(A.I. 38)}$$

Although tensors of higher ranks are hardly used in this book, we shall sometimes deal with the results of their convolution. We have seen that the differentiation of a scalar function, i.e. an invariant expression, leads to the formation of the vector gradient $\partial\Phi/\partial x_i$. The differentiation of a vector leads to the formation of a tensor of second rank $\partial^2\Phi/dx_i dx_k$. We have already seen that the convolution of this expression results in the invariant (A.I.34):

$$\frac{\partial^2\Phi}{\partial x_i^2} = \frac{\partial^2\Phi}{\partial x_i'^2}.$$

If a tensor of the second rank has the components depending on coordinates, the differentiation of these components leads to the formation of a tensor of the third rank. For example, from the tensor f_{ik} we get the tensor of the third rank $\partial f_{ik}/\partial x_l$.

Let us perform the convolution of this tensor over the indices k and l and see that it results in four quantities forming the components of a vector. Indeed,

$$\frac{\partial f_{ik}}{\partial x_k} = \frac{\partial}{\partial x'_m}(a_{is}a_{kp}f'_{sp})\frac{\partial x'_m}{\partial x_k} = a_{is}a_{kp}a_{km}\frac{\partial f'_{sp}}{\partial x'_m} =$$

$$= a_{is}a_{pm}\frac{\partial f'_{sp}}{\partial x'_m} = a_{is}\frac{\partial f'_{sm}}{\partial x'_m}. \quad \text{(A.I. 39)}$$

Since the indices k, m, s are mute, we see that the quantities $\partial f_{ik}/\partial x_k$ are transformed according to the vector transformation law (A.I.32).

Also let us consider now how to obtain an invariant expression from the components of tensors of the second rank. If the product of components of two vectors forms, as we have seen, a tensor of the second rank, it can readily be shown that the product of components of two tensors of the second rank yields a tensor of the fourth rank. Let the components of the tensor $\mathfrak{F}$ be denoted by F_{ik} and those of the tensor f by f_{lm}. Their product $T_{iklm} = F_{ik}f_{lm}$ is the tensor of the fourth rank. Let us perform the convolution of this tensor over the indices i and l as well as over k and m, i.e. let us compose the expression

$$T_{ikik} = F_{ik}f_{ik}, \qquad \text{(A.I. 40)}$$

which represents the sum of the pairwise products of the respective components. We shall make sure that this expression does not change on transition from one reference frame to another. This result is easy to prove since the transformation rule for the components F_{ik} and f_{lm} is known:

$$F_{ik}f_{lm} = \alpha_{is}\alpha_{kp}F'_{sp}\alpha_{lr}\alpha_{mt}f'_{rt}.$$

Putting $i = l$, $k = m$, we shall get

$$F_{ik}f_{ik} = \alpha_{is}\alpha_{ir}\alpha_{kp}\alpha_{kt}F'_{sp}f'_{rt} = \delta_{rs}\delta_{pt}F'_{sp}f_{rt} = F'_{sp}f'_{sp}. \qquad \text{(A.I. 41)}$$

The equality (A.I.41) provides the evidence of the invariance of (A.I.40). Of course, the invariance of F^2_{ik} or f^2_{ik} represents a special case of (A.I.40).

Since we have repeatedly made use of the Gauss-Ostrogradsky theorem to treat the vectors resulting from the three-dimensional convolution of a tensor, we shall write out the requisite formulae. In the three-dimensional space this theorem has a bearing on the transformation of a vector flow across the closed surface S to the integral of the volume $\mathscr{V}$ enveloped by that surface, e.g.,

$$\oint_S \boldsymbol{D}\, dS = \int_{\mathscr{V}} \operatorname{div} \boldsymbol{D}\, d\mathscr{V}. \qquad \text{(A.I. 42)}$$

The same theorem takes the following form in the symmetric notation:

$$\oint_S D_\alpha n_\alpha\, dS = \int_{\mathscr{V}} \frac{\partial D_\alpha}{\partial x_\alpha}\, d\mathscr{V}, \qquad \text{(A.I. 43)}$$

where n_α are the components of the normal to the surface element dS. Applying Eq. (A.I.43) to the vector

$$A_\beta = \frac{\partial f_{\alpha\beta}}{\partial x_\alpha},$$

we shall obtain

$$\int\limits_{\mathscr{V}} \frac{\partial f_{\alpha\beta}}{\partial x_\alpha} d\mathscr{V} = \oint\limits_{S} f_{\alpha\beta} n_\alpha \, dS. \tag{A.I.44}$$

§ 6. Some data on determinants. The dual tensors. 1. Let us arrange n^2 elements, denoted by the symbol a_{ik}, where i, k take on all values from 1 to n, in the form of a square table. Let the first index i in the symbol a_{ik} denote the number of the row and the second index k the number of the column of an element. Thus, we shall obtain the square matrix formed by the elements a_{ik}:

$$a_{ik} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

From this matrix we can form the determinant

$$D_n = | a_{ik} | = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix},$$

which implies a definite operation performed over the elements a_{ik}, that is the formation of the sum of $n!$ elements a_{ik}. This sum can be obtained as follows. Take the product of elements involving the elements from different rows, e.g. the rows 1 2 ... n:

$$a_{1\alpha} a_{2\beta} \dots n_{n\tau}, \tag{A.I.45}$$

or the product of elements involving the elements from different columns, e.g., 1 2 ... n:

$$a_{\alpha 1} a_{\beta 2} \dots a_{\tau n}, \tag{A.I.46}$$

where the values of the indices α, β, ..., τ will be now defined. To obtain the value of the determinant, let us compose the algebraic sum of the terms (A.I.45) or (A.I.46) differing from one another by the indices α, β, ..., τ forming a certain permutation of the nat-

ural sequence of numbers 1 2 ... n in each term of the sum This means that the indices α, β, ..., τ have different values in each term of the sum. The sum is taken over all the permutations of numbers 1 2 ... n, the total number of such permutations being equal to $n!$.

The sign "+" or "—" is ascribed to each term of the sum depending on whether an even or an odd number of pairwise permutations (transpositions) of elements is needed to obtain a given arrangement α β ... τ from the natural sequence of numbers 1 2 ... n. The pairwise transposition consists, for example, in the transition from the sequence 1 2 3 4 to the sequence 1 3 2 4 involving the transposition of the digits 2 and 3. The number of transpositions needed to accomplish the transition from the natural sequence to a given one is denoted by the letter r. So, according to the definition, a determinant of the nth order is written in full as follows:

$$D_n = |a_{ik}| = \begin{vmatrix} a_{11} & a_{12} & \dots & a_n \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \sum^{\alpha\neq\beta\neq\dots\neq\tau} (-1)^r a_{\alpha 1} a_{\beta 2} \dots a_{\tau n} =$$

$$= \sum^{\alpha\neq\beta\neq\dots\neq\tau} (-1)^r a_{1\alpha} a_{2\beta} \dots a_{n\tau}, \qquad \text{(A.I.47)}$$

where the sum is taken over all the permutations of indices α β ... τ taking on various values from 1 to n. The two last rows of the equality indicate one of the basic properties of a determinant, that is the equivalence of rows and columns. Of course, in expressions of the forms (A.I.45) and (A.I.46) it is not obligatory to take respectively the first or second indices arranged in the natural sequence. But then the bringing of specific elements to the canonical form (A.I.45) or (A.I.46) would require the indices α, β, ..., τ to be re-denoted. This new notation would be reduced to the transposition of the indices; in the determinant itself this would mean the transposition of rows or columns. Hence it is clear that the transposition of an odd number of rows (or columns) alters the sign of the determinant while the transposition of an even number of rows (or columns) leaves the value of the determinant unchanged. Having fixed a definite value n of indices i, k, ..., s, we can compose the sum of the terms of transposed indices α β .. τ, each term being accompanied with the requisite sign, i.e.

$$\sum^{\alpha\neq\beta\neq\dots\neq\tau} (-1)^r a_{\alpha i} a_{\beta k} \dots a_{\tau s},$$

whose value will be equal to $+D_n$ or $-D_n$ depending on what number of transpositions, odd or even, is needed to obtain the sequence $i\ k \ldots s$ from the natural sequence $1\ 2 \ldots n$.

This is how the determinant of the third rank D_3 is written in full:

$$D_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \sum^{\alpha \neq \beta \neq \gamma} (-1)^r a_{1\alpha} a_{2\beta} a_{3\gamma} =$$

$$= \underset{(\alpha=1,\ \beta=2\ \gamma=3)}{a_{11}a_{22}a_{33}} + \underset{(312)}{a_{13}a_{21}a_{32}} + \underset{(231)}{a_{12}a_{23}a_{31}} - \underset{(321)}{a_{13}a_{22}a_{31}} - \underset{(132)}{a_{11}a_{23}a_{32}} - \underset{(213)}{a_{12}a_{21}a_{33}}.$$

2. The determinant calculation technique. Let us pick from the sum (A.I.47) all the terms containing a certain element a_{ik}, group them together and take this element as a common factor for this combination of terms. The coefficient of the element a_{ik} thus obtained will be denoted by A_{ik} and referred to as an *adjunct* or *co-factor* of the element a_{ik}. The co-factor of a given element is calculated in accordance with a simple rule. In the determinant D_n we cross out the row and the column which contain the element a_{ik}, whose co-factor A_{ik} is sought for. Having crossed out the ith row and the kth column, we obtain the determinant D_{n-1} of the $(n-1)$th rank which is referred to as a minor Δ_{ik} of the element a_{ik}:

$$\Delta_{ik} = D_{n-1} = \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1\ k-1} & a_{1,\ k+1} & \ldots & a_{1n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{i-1,\ 1} & a_{i-1,\ 2} & \ldots & a_{i-1,\ k-1} & a_{i-1,\ k+1} & \ldots & a_{i-1,\ n} \\ a_{i+1,\ 1} & a_{i+1,\ 2} & \ldots & a_{i+1\ k-1} & a_{i+1,\ k+1} & \ldots & a_{i+1,\ n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \ldots & a_{n\ k-1} & a_{n,\ k+1} & \ldots & a_{nn} \end{vmatrix}.$$

The co-factor A_{ik} may differ from the minor Δ_{ik} only in sign:

$$A_{ik} = (-1)^{i+k} \Delta_{ik}.$$

Each element has its own corresponding co-factor, but a given element does not enter every term of the sum (A.I.47). We can, however, select a definite number of elements of a determinant which together with their co-factors permit the value of the determinant to be found. In fact, there is a theorem stating that the determinant can be expanded into the elements of any row or any column as follows:

$$D_n = \sum_{\alpha=1}^{n} a_{\alpha k} A_{\alpha k} = \sum_{\beta=1}^{n} a_{k\beta} A_{k\beta},$$

the summation over k is not carried out here while k itself may take on any value from 1 to n. If we compose the sum of the products of the elements of any row (or any column) and the minors of another row (or another column), it will be equal to zero:

$$\sum_{\alpha=1}^{n} a_{\alpha k} A_{\alpha i} = 0 \quad (\alpha \neq i).$$

The last two formulae are combined into one:

$$\sum_{\alpha=1}^{n} a_{\alpha k} A_{\alpha i} = D\delta_{ik}.$$

For example, let us calculate the determinant of the Lorentz transformation matrix, having expanded it into the elements of the first row:

$$D^{L} = \begin{vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} \end{vmatrix} = \begin{vmatrix} \Gamma & 0 & 0 & iB\Gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -iB\Gamma & 0 & 0 & \Gamma \end{vmatrix} =$$

$$= \Gamma \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \Gamma \end{vmatrix} - iB\Gamma \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -iB\Gamma & 0 & 0 \end{vmatrix} = \Gamma^2 - B^2\Gamma^2 = \Gamma^2(1 - B^2) = 1.$$

The reader can easily ascertain that the multiplication of the elements of the first row by the co-factors of the elements of other rows yields zero.

3. Let us introduce a fully antisymmetric unit tensor of the nth rank. This is the tensor $\delta_{\alpha\beta\ldots\tau}$ whose components alter sign when any two indices are interchanged and all components differing from zero are equal either to $+1$ or to -1. Any component of an antisymmetric tensor $\delta_{\alpha\beta\ldots\tau}$ whose two indices are equal turns into zero (the transposition of two such indices alters the component's sign due to the antisymmetry of the tensor, but at the same time we get the same component; thus, only zero can be equal to itself when taken with an opposite sign). Thus, only those components of the tensor $\delta_{\alpha\beta\ldots\tau}$ differ from zero whose all indices are different. Suppose $\delta_{12\ldots n} = 1$; then all the components differing from zero are equal to $+1$ if the arrangement $\alpha\beta\ldots\tau$ is obtained from the sequence $1\ 2\ \ldots\ n$ through an even number of transpositions. If the number of such transpositions in the arrangement $\alpha\beta\ldots\tau$ is odd, the component $\delta_{\alpha\beta\ldots\tau}$ is equal to -1. Making use of a fully antisymmetric unit tensor, we can rewrite the expres-

sion for the determinant D_n as follows:

$$D_n = \delta_{\alpha\beta \ldots \tau} a_{1\alpha} a_{2\beta} \ldots a_{n\tau} = \delta_{\alpha\beta \ldots \tau} a_{\alpha 1} a_{\beta 2} \ldots a_{\tau n},$$

where the summation over the pairs of indices $\alpha\beta \ldots \tau$ is implied this time.

In particular, we can write the determinant corresponding to the Lorentz matrix:

$$D^{L} = \delta_{\alpha\beta\gamma\rho} \alpha_{1\alpha} \alpha_{2\beta} \alpha_{3\gamma} \alpha_{4\rho} = 1.$$

4. Now we shall deal with the 4-space of the STR. First of all, it should be pointed out that we have defined a fully antisymmetric unit tensor $\delta_{\beta\gamma\rho\mu}$ without proving it to be a tensor. We must make sure that the components of this tensor have the same values in all IFRs, i.e. under the Lorentz transformation. This can be done quite easily. In accordance with the tensor component transformation rule

$$\delta'_{iklm} = \alpha_{i\beta} \alpha_{k\gamma} \alpha_{l\rho} \alpha_{m\mu} \delta_{\beta\gamma\rho\mu}.$$

But according to what was said in item 1 of this section the quantity on the right-hand side is equal to $D^{L}\delta_{iklm}$, i.e. is equal to ± 1 depending on the number of transpositions needed to obtain the arrangement $iklm$ from the natural one. This means that $\delta_{\beta\gamma\rho\mu}$ has the same components in any IFR. The components of this tensor also do not change on transition from the left system of coordinates to the right one (i.e. when one or three spatial coordinates change sign). According to Eq. (A.I.30) the components of this tensor had to alter sign in this case. Consequently, $\delta_{\beta\gamma\rho\mu}$ is not a tensor but a pseudotensor; its components behave in a different way, as compared to tensors, when a coordinate alters sign (is reflected). In all other transformations the behaviour of these components coincides with that of the components of a tensor.

5. The cross and the mixed product of vectors in a three-dimensional space. These questions are treated in order to have a good analogy when considering some of the quantities in the 4-space of the STR.

Consider three unit vectors $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, $\boldsymbol{m}_3$ of the orthogonal Cartesian system of coordinates. Compose the cross product of any pair of these vectors $[\boldsymbol{m}_\alpha \boldsymbol{m}_\beta]$; as a result, the third vector will be obtained with the "plus" or "minus" sign, depending on the order of co-factors in the cross product. This cross product can be easily written via fully antisymmetric unit tensor of the third rank:

$$[\boldsymbol{m}_\alpha \boldsymbol{m}_\beta] = \delta_{\alpha\beta\gamma} \boldsymbol{m}_\gamma.$$

Now we can write the cross product of the vectors $a_1 = a_{1\alpha} m_\alpha$ and $a_2 = a_{2\beta} m_\beta$:

$$[a_1 a_2] = [a_{1\alpha} m_\alpha,\ a_{2\beta} m_\beta] = a_{1\alpha} a_{2\beta} [m_\alpha m_\beta] = \delta_{\alpha\beta\gamma} a_{1\alpha} a_{2\beta} m_\gamma. \qquad (A.I.48)$$

It is seen from Eq. (A.I.48) that m_γ has the coefficients formed by the products of vector components and convoluted with the tensor $\delta_{\alpha\beta\gamma}$. Let us rewrite Eq. (A.I.48):

$$[a_1 a_2] = \delta_{\alpha\beta\gamma} a_{1\alpha} a_{2\beta} m_\gamma = \frac{1}{2} (\delta_{\alpha\beta\gamma} a_{1\alpha} a_{2\beta} + \delta_{\beta\alpha\gamma} a_{1\beta} a_{2\alpha}) m_\gamma =$$

$$= \frac{1}{2} \delta_{\alpha\beta\gamma} (a_{1\alpha} a_{2\beta} - a_{2\alpha} a_{1\beta}) m_\gamma = \frac{1}{2} \delta_{\alpha\beta\gamma} C_{\alpha\beta} m_\gamma.$$

In the third link of this equality a second term is added which is equal to the first term with the mute indices α and β interchanged. Here in the third link we take into account that $\delta_{\beta\alpha\gamma} = -\delta_{\alpha\beta\gamma}$. In the fourth link $\delta_{\alpha\beta\gamma}$ is taken out of the parentheses. The antisymmetric tensor, thus formed in the parentheses, is denoted by $C_{\alpha\beta} = a_{1\alpha} a_{2\beta} - a_{1\beta} a_{2\alpha}$ Consequently, the cross product $[a_1 a_2]$ is a vector whose components are obtained from the antisymmetric tensor $C_{\alpha\beta}$ according to the formulae

$$C_\gamma = \frac{1}{2} \delta_{\alpha\beta\gamma} C_{\alpha\beta}.$$

The vector $C(C_\gamma)$ is said to be dual to the antisymmetric tensor $C_{\alpha\beta}$. This means that the vector C is orthogonal to two vectors a_1 and a_2 defining a two-dimensional plane. The orthogonality can be proved analytically at once:

$$C a_1 = C_\gamma a_{1\gamma} = \frac{1}{2} \delta_{\alpha\beta\gamma} C_{\alpha\beta} a_{1\gamma} = \frac{1}{2} \delta_{\alpha\beta\gamma} \{a_{1\alpha} a_{2\beta} a_{1\gamma} - a_{1\beta} a_{2\alpha} a_{1\gamma}\} = 0.$$

This expression is equal to zero since both the minuend and the subtrahend are the determinants with two equal rows and such determinants are equal to zero. In much the same way it is proved that $C a_2 = 0$. In geometrical terms the norm of the vector C is equal to the area of a parallelogram constructed on the vectors a_1 and a_2.

The mixed product of three vectors a_1, a_2, a_3 is denoted by (a_1, a_2, a_3) and defined as follows:

$$(a_1,\ a_2,\ a_3) = a_1 [a_2 a_3] = a_{1\lambda} m_\lambda \delta_{\alpha\beta\gamma} a_{2\alpha} a_{3\beta} m_\gamma = \delta_{\alpha\beta\gamma} a_{1\lambda} a_{2\alpha} a_{3\beta} m_\lambda m_\gamma =$$

$$= a_{1\lambda} \delta_{\lambda\gamma} a_{2\alpha} \delta_{\alpha\beta\gamma} = \delta_{\alpha\beta\gamma} a_{1\gamma} a_{2\alpha} a_{3\beta} = D_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Here $\delta_{\lambda\gamma}$ is the Kronecker delta (A.I.4).

In geometrical terms the mixed product of three vectors defines the volume of the parallelepiped constructed on these vectors. This volume is obtained with the "+" or "—" sign depending on the order of the vectors a_1, a_2, a_3 in the mixed product.

6. The dual tensors. Let two 4-vectors a_1 and a_2 be given in the 4-space. Then the projections of the parallelogram area on the coordinate planes (x_i, x_k) are defined by the antisymmetric tensor $\xi_{ik} = a_{1i}a_{2k} - a_{2i}a_{ik}$. In the 4-space each area element ξ_{ik} can be brought into correspondence with another normal area element ξ^*_{ik} such that all the straight lines lying in it are normal to all the straight lines of the initial area element. If the element ξ^*_{ik} orthogonal to ξ_{ik} has the same area as ξ_{ik}, the element ξ^*_{ik} is called dual with respect to ξ_{ik}. It can be shown that

$$\xi^*_{ik} = \frac{1}{2}\delta_{iklm}\xi_{lm}. \qquad \text{(A.I.49)}$$

With the aid of this formula any antisymmetric tensor can be brought into correspondence with its dual tensor. The dual tensor f^*_{ik} is, in a certain sense, equivalent to the initial tensor f_{ik}. We have seen that the second group of Maxwell's equations takes the simpler form when written via the dual tensor F^*_{ik}. The sum of the products of the antisymmetric tensor components by their dual co-factors yields a pseudoscalar

$$F_{ik}F^*_{ik} = \frac{1}{2}\delta_{iklm}F_{ik}F_{lm} = \frac{1}{2}\alpha_{ia}\alpha_{kb}\alpha_{lc}\alpha_{md}\delta_{abcd}\alpha_{ir}\alpha_{ks}F'_{rs}\alpha_{lt}\alpha_{mn}F'_{tn} =$$
$$= \frac{1}{2}\delta_{ar}\delta_{bs}\delta_{ct}\delta_{dn}\delta_{abcd}F'_{rs}F'_{tn} = \frac{1}{2}\delta_{rstn}F'_{rs}F'_{tn} = F'_{rs}F'^*_{rs}.$$

In these equalities we made use of the definition (A.I.49), the tensor transformation formulae (A.I.30) and the properties of the Lorentz matrix coefficients (A.I.15). It can be easily shown that $F^2_{ik} = F_{ik}F_{ik}$ is an invariant:

$$F_{ik}F_{ik} = \alpha_{ia}\alpha_{kb}F'_{ab}\alpha_{ic}\alpha_{kd}F'_{cd} = \delta_{ac}\delta_{bd}F'_{ab}F'_{cd} = F'_{ab}F'_{ab}.$$

These two invariants were used in § 6.5.

§ 7. **The stress tensor.** The stress tensor is introduced in continuum mechanics to characterize the force acting on a volume as a whole via the force acting on the surface confining this volume. We obtained the expressions of this type studying the forces in an electromagnetic field; it is useful to consider this problem in mechanics where the physical essence of phenomena is most obvious.

If an elastic body is subjected to deformation forces arise in it tending to return it to the state of equilibrium. These forces are

referred to as the internal stresses and are caused by the forces of interaction of the molecules of the body. The distinctive feature of these forces is their "small radius of action"; in other words, their influence is felt only over microscopic (atomic) distances. Therefore, it is clear that when a certain volume $\mathscr{V}$ is considered inside a body, the forces acting on that volume reduce to the forces acting across the surface confining that volume.

Indeed, let the force $\boldsymbol{F}$ act on a unit of volume of a body. Let us single out the volume $\mathscr{V}$ within a body and consider the total force acting on this volume. If the volume $d\mathscr{V}$ is subjected to the force $\boldsymbol{F}\,d\mathscr{V}$, the total force acting on the volume is equal to

$$\int_{\mathscr{V}} \boldsymbol{F}\,d\mathscr{V}. \tag{A.I.50}$$

The various parts of the considered volume interact, but in accordance with the law of equal action and reaction the forces of interaction cancel out and yield a zero resultant. Thus, the total force acting on the volume $\mathscr{V}$ arises due to the forces exerted by the surrounding parts of the body. But as we have already mentioned, these forces act only across the surface confining the considered volume. Therefore, the total force will reduce to a certain surface integral. In particular, the βth component of the force

$$\int_{\mathscr{V}} F_\beta\,d\mathscr{V} \tag{A.I.51}$$

must also turn into a surface integral. However, this is possible only if F_β can be represented in the form

$$F_\beta = \frac{\partial T_{\alpha\beta}}{\partial x_\beta}, \tag{A.I.52}$$

where $T_{\alpha\beta}$ denotes the components of a tensor (only in this case the convolution results in a vector). Then in accordance with Eq. (A.I.44)

$$\int_{\mathscr{V}} F_\beta\,d\mathscr{V} = \int_{\mathscr{V}} \frac{\partial T_{\alpha\beta}}{\partial x_\beta}\,d\mathscr{V} = \oint_S T_{\alpha\beta} n_\alpha\,dS. \tag{A.I.53}$$

Having multiplied both sides of Eq. (A.I.53) by $\boldsymbol{m}_\beta$ and carried out the actual summation, we obtain

$$\int_{\mathscr{V}} \boldsymbol{F}\,d\mathscr{V} = \int_{\mathscr{V}} \frac{\partial T_{\alpha\beta}}{\partial x_\alpha}\,\boldsymbol{m}_\beta\,d\mathscr{V} = \oint_S T_{\alpha\beta} n_\alpha \boldsymbol{m}_\beta\,dS. \tag{A.I.54}$$

The relation (A.I.54) shows that the total force acting on the volume is reduced to the surface integral. Consequently, the result

obtained can be formulated as follows: if the force $\boldsymbol{F}$ acting on a unit of volume can be represented as

$$\boldsymbol{F} = \frac{\partial T_{\alpha\beta}}{\partial x_{\beta}} \boldsymbol{m}_{\beta}, \tag{A.I.55}$$

its action on the whole volume can be described as the action of a surface force distributed over the surface confining this volume,

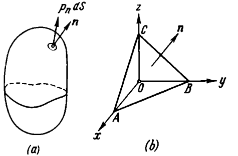


Fig. A.2. (*a*) The stress exerted on the area dS at the boundary surface of the volume in which the stresses are generated; $\boldsymbol{n}$ is the normal to the surface element dS; $\boldsymbol{p}_n$ is the force acting on the area whose normal is $\boldsymbol{n}$. (*b*) On derivation of the condition for the equilibrium of a closed tetrahedron-shaped elementary volume. The outward normal is chosen as a normal to the closed tetrahedron surface. At the faces *BOC*, *AOC* and *AOB* the unit vectors of the normal are equal to $-\boldsymbol{i}$, $-\boldsymbol{j}$ and $-\boldsymbol{k}$ respectively. The areas of the faces *BOC*, *AOC* and *AOB* are equal to $dS\cos(\widehat{\boldsymbol{n}, x})$, $dS\cos(\widehat{\boldsymbol{n}, y})$ and $dS\cos(\widehat{\boldsymbol{n}, z})$, respectively.

with the surface element dS, whose normal $\boldsymbol{n}$ has the unit vector components n_{α}, being subjected to the force

$$T_{\alpha\beta} n_{\alpha} \boldsymbol{m}_{\beta}. \tag{A.I.56}$$

Let us discuss briefly the physical meaning of the stress tensor components. Let us return to the volume $\mathscr{V}$ within the body experiencing a deformation. The force acting on the surface element dS confining the volume $\mathscr{V}$ depends on the value and direction of the element dS, i.e. on the direction of the normal $\boldsymbol{n}$ relative to this element. Let us denote this force by $\boldsymbol{p}_n\,dS$, having pointed out that its direction, generally speaking, does not coincide with the direction of the normal of the surface element dS (Fig. A.2*a*). The vector $\boldsymbol{p}_n$ is the force per unit of area; it depends on the orientation of the surface element and is called the stress on the surface element dS with the normal $\boldsymbol{n}$. At each point of the deformed elastic body any direction $\boldsymbol{n}$ has its corresponding stress vector $\boldsymbol{p}_n$. In each Cartesian reference frame it is possible to determine the stresses $\boldsymbol{p}_x$, $\boldsymbol{p}_y$, $\boldsymbol{p}_z$ acting on unit of area elements whose normals

coincide with the coordinate axes. We shall demonstrate that the stress relating to any area element dS with the given vector $\boldsymbol{n}$ can be expressed through nine components of the vectors $\boldsymbol{p}_x$, $\boldsymbol{p}_y$, $\boldsymbol{p}_z$, these nine components taken together form the stress tensor $T_{\alpha\beta}$.

Let an elastically deformed body be in equilibrium. We shall consider the infinitesimal tetrahedron $OABC$ (Fig. A.2b) whose inclined side has the area dS. Let the normal $\boldsymbol{n}$ of this side be directed at an acute angle to the x axis. Then the area elements cut out by the coordinate planes will equal $dS\cos(\widehat{\boldsymbol{n}, x})$, $dS\cos(\widehat{\boldsymbol{n}, y})$ and $dS\cos(\widehat{\boldsymbol{n}, z})$. The normals of these area elements are oriented in the direction opposite to the unitary coordinate axes $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$, so that the side BOC is subjected to the force $-\boldsymbol{p}_x\,dS\cos(\widehat{\boldsymbol{n}, x})$. The value of $\boldsymbol{p}_x$ can be taken at any point of BOC since this side is infinitesimal. In much the same manner, the forces acting on the sides AOC and AOB turn out to be equal to $-\boldsymbol{p}_y\,dS\cos(\widehat{\boldsymbol{n}, y})$ and $-\boldsymbol{p}_z\,dS\cos(\widehat{\boldsymbol{n}, z})$ respectively. In equilibrium the resultant of the forces acting on the tetrahedron is equal to zero

$$dS\,[\boldsymbol{p}_n - \boldsymbol{p}_x\cos(\widehat{\boldsymbol{n}, x}) - \boldsymbol{p}_y\cos(\widehat{\boldsymbol{n}, y}) - \boldsymbol{p}_z\cos(\widehat{\boldsymbol{n}, z})] = 0,$$

whence the sought for stress $\boldsymbol{p}_n$ is expressed via $\boldsymbol{p}_x$, $\boldsymbol{p}_y$, $\boldsymbol{p}_z$:

$$\boldsymbol{p}_n = \boldsymbol{p}_x\cos(\widehat{\boldsymbol{n}, x}) + \boldsymbol{p}_y\cos(\widehat{\boldsymbol{n}, y}) + \boldsymbol{p}_z\cos(\widehat{\boldsymbol{n}, z}). \quad \text{(A.I.57)}$$

Passing to the symmetric notation in Eq. (A.I.57), we can show that we obtained a tensor. Indeed, $\boldsymbol{n}$ is the vector of the normal of an arbitrary side with the components n_α and therefore, $\boldsymbol{p}_n = \boldsymbol{p}_1 n_1 + \boldsymbol{p}_2 n_2 + \boldsymbol{p}_3 n_3 = \boldsymbol{p}_\alpha n_\alpha$. But $\boldsymbol{p}_\alpha = p_{\alpha\beta}\boldsymbol{m}_\beta$ where $p_{\alpha\beta}$ are the components of the vector $\boldsymbol{p}_\alpha$, and consequently,

$$\boldsymbol{p}_n = p_{\alpha\beta}\boldsymbol{m}_\beta n_\alpha. \quad \text{(A.I.58)}$$

It is seen from Eq. (A.I.58) that the nine components of the vectors $\boldsymbol{p}_\alpha$ are transformed as a tensor (cf., e.g., Eq. (A.I.24)).

§ 8. The rectilinear oblique-angled systems of coordinates. Until now we have used the orthogonal rectilinear system of coordinates, but a transition to rectilinear oblique-angled systems of coordinates makes it possible to illustrate the features characteristic of arbitrary coordinate systems dealt with in this book.

Let us choose, as before, a set of straight lines, not orthogonal this time, as coordinate axes and denote them by x^1 and x^2 (Fig. A.3a). Then let us mark a basis unit vector $\boldsymbol{m}_\mu$ on each of these axes x^μ.

An arbitrary vector $\boldsymbol{A}$ can be expanded into the non-colinear vectors $\boldsymbol{m}_\mu$:

$$\boldsymbol{A} = A^\mu \boldsymbol{m}_\mu. \quad \text{(A.I.59)}$$

The quantities A^μ are the components of the vector $\boldsymbol{A}$ which are obtained by means of the parallel projection of this vector on the coordinate axes; according to the definition they are called the *contravariant* components of the vector $\boldsymbol{A}$.

The quantities

$$A_\mu = \boldsymbol{A}\boldsymbol{m}_\mu \qquad \text{(A.I.60)}$$

are the orthogonal projections of the vector $\boldsymbol{A}$ on the coordinate axes and are referred to as the *covariant* components of the vec-

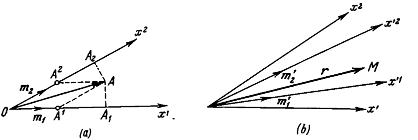


Fig. A.3. The definiton of the co- and contravariant coordinates in a rectilinear oblique-angled system of coordinates on the plane.

tor $\boldsymbol{A}$. Obviously, these definitions can be retained for any number of measurements. Designating the scalar product of the basis vectors

$$\boldsymbol{m}_\mu \boldsymbol{m}_\nu = g_{\mu\nu}, \qquad \text{(A.I.61)}$$

we get $g_{\mu\nu} = g_{\nu\mu}$ and in the case of the rectilinear coordinate axes $g_{\mu\nu} = \text{const}$. The covariant and contravariant coordinates relate to the same vector and are interrelated:

$$A_\mu = \boldsymbol{A}\boldsymbol{m}_\mu = A^\nu \boldsymbol{m}_\nu \boldsymbol{m}_\mu = g_{\mu\nu} A^\nu. \qquad \text{(A.I.62)}$$

This equation defines the transition from the contravariant components of a vector to the covariant components.

Then, let us define the quantities $g^{\mu\nu}$ by the condition:

$$g_{\mu\rho} g^{\nu\rho} = \delta^\nu_\mu, \quad \delta^\nu_\mu = \begin{cases} 0, & \nu \neq \mu, \\ 1, & \nu = \mu. \end{cases} \qquad \text{A.I.63)}$$

Now let us construct the expression

$$g^{\mu\rho} A_\rho = g^{\mu\rho} g_{\rho\sigma} A^\sigma = \delta^\mu_\sigma A^\sigma = A^\mu. \qquad \text{(A.I.64)}$$

The last equality defines the transition from the covariant to contravariant components. Thus, we have obtained two fundamental formulae of transition:

$$A_\mu = g_{\mu\nu} A^\nu, \quad A^\mu = g^{\mu\nu} A_\nu. \qquad \text{(A.I.65)}$$

The determinant formed by the quantities g_{ik} is denoted by g (see § 6 of this Appendix):

$$g = |g_{ik}|.$$

Using the formula (p. 380)

$$\sum g_{\alpha k} A_{\alpha l} = \delta_{jk} g$$

we immediately obtain

$$g^{\mu\nu} = \frac{A_{\mu\nu}}{g}, \qquad \text{(A.I.66)}$$

where $A_{\mu\nu}$ is the cofactor of the element $g_{\mu\nu}$.

It is easily seen that for the orthogonal rectilinear coordinates, when $\boldsymbol{m}_\mu \boldsymbol{m}_\nu = \delta_{\mu\nu}$, $g_{\mu\nu} = \delta_{\mu\nu}$ and $A_\mu = A^\mu$, i.e. there is no difference between the covariant and contravariant coordinates. That is why in the case of the orthogonal Cartesian system of coordinates we simply speak of the coordinates of vectors.

By the definition the scalar product of two vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ is the quantity

$$\boldsymbol{AB} = (A^\mu \boldsymbol{m}_\mu)(B^\nu \boldsymbol{m}_\nu) = A^\mu B^\nu (\boldsymbol{m}_\mu \boldsymbol{m}_\nu) = g_{\mu\nu} A^\mu B^\nu = A_\nu B^\nu. \qquad \text{(A.I.67)}$$

The scalar product of a vector by itself defines the square of the vector's absolute value or the norm of the vector:

$$A^2 = g_{\mu\nu} A^\mu A^\nu = A_\nu A^\nu. \qquad \text{(A.I.68)}$$

Thus, the norm is the square of the vector's length. If the norm of the vector is equal to unity, the vector is called normed or unitary. If the norm of any non-zero vector is positive, the space is referred to as the *proper Euclidean space*.

In particular, the square of the infinitesimal vector $d\boldsymbol{r}$ possessing the components dx^ν and connecting two infinitely close points of space is equal to

$$d\boldsymbol{r}^2 = ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\mu. \qquad \text{(A.I.69)}$$

Let us change the system of base axes and pass to new basis vectors $\boldsymbol{m}'_\mu$ directed along the rectilinear axes x'^μ (Fig. A.3*b*). Any new vector $\boldsymbol{m}'_\mu$ can be expanded into the old basis vectors:

$$\boldsymbol{m}'_\mu = a^\nu_\mu \boldsymbol{m}_\nu, \qquad \text{(A.I.70)}$$

where a^ν_μ are constant coefficients dependent on the concrete transformation of oblique-angled axes. The quantity $\boldsymbol{m}'_\mu$ does not change under any transformation provided $|a^\nu_\mu| = a \neq 0$. Of course, any old vector $\boldsymbol{m}_\mu$ can be expanded into the new ones:

$$\boldsymbol{m}_\nu = a'^\lambda_\mu \boldsymbol{m}'_\lambda, \quad a' = |a'^\lambda_\nu| \neq 0. \qquad \text{(A.I.71)}$$

It follows from Eqs. (A.I.70) and (A.I.71) that

$$m'_\mu = a^\nu_\mu m_\nu = a^\nu_\mu a'^\lambda_\nu m'_\lambda; \quad m_\lambda = a'^\lambda_\nu m'_\lambda = a'^\lambda_\nu a^\mu_\lambda m_\mu. \qquad (A.I.72)$$

From Eq. (A.I.72) the coefficients a^μ_ν and a'^μ_ν are seen to be related by the following expressions:

$$a^\nu_\mu a'^\lambda_\nu = \delta^\lambda_\mu, \quad a'^\lambda_\nu a^\mu_\lambda = \delta^\mu_\nu, \qquad (A.I.73)$$

where δ^λ_μ is defined in accordance with Eq. (A.I. 63).

The radius vector $\boldsymbol{r}$ drawn from the origin of coordinates to the point M (Fig. A.3*b*) can be written in two forms:

$$x^\nu m_\nu = x'^\mu m'_\mu. \qquad (A.I.74)$$

Taking into account Eqs. (A.I.70) and (A.I.71), the last equation can be rewritten in two forms as well:

$$x^\nu a'^\mu_\nu m'_\mu = x'^\mu m'_\mu, \quad x^\nu m_\nu = x'^\mu a^\nu_\mu m_\nu, \qquad (A.I.75)$$

whence follow the formulae for the direct and inverse transformations of the contravariant coordinates of the vector $\boldsymbol{r}$:

$$x'^\mu = a'^\mu_\nu x^\nu; \quad x^\nu = a^\nu_\mu x'^\mu. \qquad (A.I.76)$$

Here is the definition of a vector: the vector $\boldsymbol{A}$ is the quantity whose covariant components are transformed under a transition to a new reference frame in the same manner as the basis vectors $\boldsymbol{m}_\mu$. The contravariant components of vectors are transformed as the contravariant coordinates x^μ. Let us find the formulae for the transformation of the components of the vector $\boldsymbol{A}$. For covariant components

$$A_\mu = \boldsymbol{A} m_\mu = \boldsymbol{A} a'^\lambda_\mu m'_\lambda = a'^\lambda_\mu A'_\lambda. \qquad (A.I.77)$$

The inverse transformation formula takes the form

$$A'_\mu = \boldsymbol{A} m'_\mu = \boldsymbol{A} a^\nu_\mu m_\nu = a^\nu_\mu A_\nu. \qquad (A.I.78)$$

On the other hand, precisely as for the vector $\boldsymbol{r}$ we can write

$$\boldsymbol{A} = A^\nu m_\nu = A'^\mu m'_\mu, \qquad (A.I.79)$$

whence

$$A^\nu a'^\mu_\nu m'_\mu = A'^\mu m'_\mu, \quad A^\nu m_\nu = A'^\mu a^\nu_\mu m_\nu \qquad (A.I.80)$$

and consequently

$$A'^\mu = a'^\mu_\nu A^\nu, \quad A^\nu = a^\nu_\mu A'^\mu. \qquad (A.I.81)$$

We see that the transformation formulae for the covariant and contravariant components of a vector are different.

Let us write the transformation law for the quantity (A.I.61).

$$g'_{\mu\nu} = (m'_{\mu} m'_{\nu}) = a^{\rho}_{\mu} m_{\rho} a^{\sigma}_{\nu} m_{\sigma} = a^{\rho}_{\mu} a^{\sigma}_{\nu} g_{\rho\sigma}. \qquad \text{(A.I.82)}$$

According to the definition, this is the transformation law for a covariant tensor. The quantity retaining its value when the basis vectors (A.I.70) and (A.I.71) change is referred to as an invariant. In the considered case of the rectilinear oblique-angled axes a^{ν}_{μ} and a'^{ν}_{μ} are constant values. Let us demonstrate the invariance of the distance between points:

$$ds'^2 = g'_{\mu\nu}\, dx'^{\mu}\, dx'^{\nu} = a^{\rho}_{\mu}\, dx'^{\mu} a^{\sigma}_{\nu}\, dx'^{\nu}\, g_{\rho\sigma} = g_{\rho\sigma}\, dx^{\rho}\, dx^{\sigma} \qquad \text{(A.I.83)}$$

(see Eq. (A.I.81)).

The invariance of the operator $\Delta = \frac{\partial^2}{\partial x_{\mu}\, \partial x^{\mu}}$ can also be easily verified.

When, for diverse reasons, vectors are introduced, both covariant and contravariant components may be found among the components of the vectors. We shall quote two important examples. From Eq. (A.I.81) it follows that

$$dA^{\nu} = a^{\nu}_{\mu}\, dA'^{\mu}; \qquad \text{(A.I.84)}$$

whence it is clear that the differentials of the contravariant coordinates of a vector are transformed as contravariant vectors. However, having considered the scalar function of contravariant components $\varphi(x^{\alpha})$ and in particular the components of the vector $\partial\varphi/\partial x^{\alpha}$, we immediately make sure that we deal with the covariant components.

Indeed, $\varphi = \varphi(x'^{\nu}) = \varphi[x'^{\nu}(x^{\alpha})]$; the coordinate transformation is implied to be known; as usual, we shall write out the formulae with "convenient" indices. Thus, in accordance with Eq. (A.I.76) $x'^{\nu} = a'^{\nu}_{\alpha} x^{\alpha}$, whence $\partial x'^{\nu}/\partial x^{\alpha} = a'^{\nu}_{\alpha}$. In accordance with the formulae for differentiating a composite function

$$\frac{\partial\varphi}{\partial x^{\alpha}} = \frac{\partial\varphi}{\partial x'^{\nu}} \frac{\partial x'^{\nu}}{\partial x^{\alpha}} = a'^{\nu}_{\alpha} \frac{\partial\varphi}{\partial x'^{\nu}}, \qquad \text{(A.I.85)}$$

this is precisely the transformation formula for the components of the covariant vector (A. I. 77). Once again we point out that all formulae obtained are valid for the space of any number of dimensions.

Passing to the 4-space-time of Minkowski, we shall recall that the consequence of two Einstein's postulates is the invariance of the quadratic form

$$ds^2 = c^2\, dt^2 - dx^2 - dy^2 - ds^2 \qquad \text{(A.I.86)}$$

on transition from one IFR to another, i.e. under the Lorentz transformation. The expression (A.I. 86) defines the square of the elementary "distance" in 4-space. But the square of the distance (A.I.86) is not necessarily positive. In this connection the Euclidean space determined by the form (A.I.80) is referred to as the *improper Euclidean* or *pseudo-Euclidean space*. To take advantage of the formalism of the proper Euclidean space, we can resort to the technique used in this book and consisting in the introduction of the imaginary coordinate (cf. Chapter 3). This technique simplifies the presentation, but at the same time it unintentionally insinuates the idea of an imaginary nature of relativistic laws themselves which, of course, have nothing to do with the number i utilized only for the purpose of making calculations easier.

Resorting to the real coordinates $x^0 = ct$, $x^1 = x$, $x^2 = y$, $x^3 = z$, we can write Eq. (A.I.86) as

$$ds^2 = dx^{0^2} - dx^{1^2} - dx^2 - dx^{3^2}. \qquad \text{(A.I.87)}$$

In any case

$$d\vec{R} = m_0\, dx^0 + m_1\, dx^1 + m_2\, dx^2 + m_3\, dx^3. \qquad \text{(A.I.88)}$$

The expression for ds^2 from Eq. (A.I.87) coincides with that from Eq. (A.I.86), provided the following conditions are satisfied:

$$m_0^2 = 1, \quad m_1^2 = m_2^2 = m_3^2 = -1; \qquad \text{(A.I.89)}$$

$$m_i m_k = 0 \quad \text{for} \quad i,\ k = 0,\ 1,\ 2,\ 3. \qquad \text{(A.I.90)}$$

All these conditions can be expressed by one formula

$$m_i m_k = g_{ik}, \qquad \text{(A.I.91)}$$

where

$$g_{ik} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \qquad \text{(A.I.92)}$$

Now g_{ik} defines the metric tensor of the pseudo-Euclidean space. Since $A_\mu = g_{\mu\nu}A^\nu$, the relationship between the covariant and contravariant vector components follows immediately from this:

$$A_0 = A^0,\ A_1 = -A^1,\ A_2 = -A^2,\ A_3 = -A^3. \qquad \text{(A.I.93)}$$

The scalar product of two vectors and the norm of a vector are defined by the following expressions:

$$AB = g_{\mu\nu}A^\mu B^\nu = A^0B^0 - A^1B^1 - A^2B^2 - A^3B^3, \qquad \text{(A.I.94)}$$

$$|A^2| = g_{\mu\nu}A^\mu A^\nu = (A^0)^2 - (A^1)^2 - (A^2)^2 - (A^3)^2. \qquad \text{(A.I.95)}$$

It is seen from Eq. (A.I.95) that the norm of an arbitrary nonzero real vector is not necessarily positive: it can also be zero or negative. This indicates once more that the four-dimensional space of the special theory of relativity is pseudo-Euclidean.

§ 9. **The definition of hyperbolic functions and some relationships between them.** For real x the basic definitions are as follows:

$$\sinh x = \frac{e^x - e^{-x}}{2}, \quad \cosh x = \frac{e^x + e^{-x}}{2},$$

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{1 - e^{-2x}}{1 + e^{2x}}.$$

It is seen at once that

$$\cosh^2 x - \sinh^2 x = \frac{1}{4}\{(e^x + e^{-x})^2 - (e^x - e^{-x})^2\} = \frac{1}{4} \cdot 4 = 1.$$

Dividing the left-hand and the extreme right-hand sides of the last equality by $\cosh^2 x$, we get

$$1 - \tanh^2 x = \frac{1}{\cosh^2 x}, \quad \text{or} \quad \cosh x = \frac{1}{\sqrt{1 - \tanh^2 x}}.$$

To obtain the formulae

$$\cosh(x_1 + x_2) = \cosh x_1 \cosh x_2 + \sinh x_1 \sinh x_2,$$
$$\sinh(x_1 + x_2) = \sinh x_1 \cosh x_2 + \sinh x_2 \cosh x_1,$$

one needs to substitute into the definitions

$$\cosh(x_1 + x_2) = \frac{e^{x_1}e^{x_2} + e^{-x_1}e^{-x_2}}{2}, \quad \sinh(x_1 + x_2) = \frac{e^{x_1}e^{x_2} - e^{-x_1}e^{-x_2}}{2}$$

the values following from the definition of hyperbolic functions:

$$e^{\theta_{1,2}} = \cosh\theta_{1,2} + \sinh\theta_{1,2}, \quad e^{-\theta_{1,2}} = \cosh\theta_{1,2} - \sinh\theta_{1,2}.$$

Finally, here is the very important formula for the real x:

$$\tanh(x_1 + x_2) = \frac{\sinh(x_1 + x_2)}{\cosh(x_1 + x_2)} =$$

$$= \frac{\sinh x_1 \cosh x_2 + \cosh x_1 \sinh x_2}{\cosh x_1 \cosh x_2 + \sinh x_1 \sinh x_2} = \frac{\tanh x_1 \tanh x_2}{1 + \tanh x_1 \tanh x_2}.$$

In this book (Chapter 5) we also make use of the expansion

$$\cosh x = \frac{e^2 + e^{-x}}{2} \approx \frac{1 + x + \frac{x^2}{2} \ \dots \ + 1 - x + \frac{x^2}{2} \ \dots}{2} =$$

$$= 1 + \frac{x^2}{2} + \ \dots, \quad \sinh x \approx x.$$

The relationship between hyperbolic and trigonometric functions is determined as follows. In the definitions of trigonometric functions (the Euler formulae)

$$\sin z = \frac{e^{iz} - e^{-iz}}{2}, \quad \cos z = \frac{e^{iz} + e^{-iz}}{2}, \quad \tan z = \frac{1}{i}\frac{e^{iz} - e^{-iz}}{e^{iz} + e^{-iz}}$$

we shall put the imaginary value of z, i.e. assume $z = i\varphi$. Then it is immediately seen that

$$\sin(i\varphi) = i \sinh \varphi, \quad \cos(i\varphi) = \cosh \varphi, \quad \tan(i\varphi) = i \tanh \varphi,$$

recalling that $i^2 = -1$ and $1/i = -i$.

BIBLIOGRAPHY TO APPENDIX I

1. McConnell, A. J. *Application of Tensor Analysis*, N. Y. 1957.

This rationally compiled book contains all the necessary information on tensor analysis utilized in physics. Having studied the book, the reader becomes quite prepared mathematically for reading books devoted to the Riemannian geometry and the general theory of relativity.

2. Kochin, N. E. *Vector Calculus and Elements of Tensor Calculus*, Moscow, 1965 (in Russian).

The book presents the basic information on tensors in very easy terms.

APPENDIX II

THE BASIC FORMULAE OF ELECTRODYNAMICS IN THE GAUSSIAN SYSTEM

In electrodynamics the Gaussian system is used as often as the SI system to which we kept in Chapter 6. Surely, the choice of the system of units does not affect the essence of the matter, but the appearance of formulae varies. For the readers' convenience we shall cite the basic formulae in the Gaussian system. The formulae possess the same numbering under which they are given in Chapter 6.

The equations for the potentials $\boldsymbol{A}$ and φ have the following form *in vacuo*:

$$\Box \boldsymbol{A} \equiv \Delta \boldsymbol{A} - \frac{1}{c^2}\frac{\partial^2 \boldsymbol{A}}{\partial t^2} = -\frac{4\pi}{c}\boldsymbol{j}, \quad \Box \varphi \equiv \Delta\varphi - \frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} = -4\pi\rho. \tag{6.9}$$

The Lorentz condition:

$$\operatorname{div} \boldsymbol{A} + \frac{1}{c}\frac{\partial\varphi}{\partial t} = 0. \tag{6.8}$$

The charge conservation law remains invariable:

$$\frac{\partial\rho}{\partial t} + \operatorname{div}\boldsymbol{j} = 0. \tag{6.4}$$

The definition of the 4-potential and 4-current:

$$\vec{\Phi}(\boldsymbol{A},\ i\varphi), \quad \vec{\Phi}(\varphi,\ \boldsymbol{A}), \tag{6.11}$$

$$\vec{s}(\boldsymbol{j},\ ic\rho), \quad \vec{s}(c\rho,\ \boldsymbol{j}). \tag{6.12}$$

The relation between the average fields and potentials:

$$\boldsymbol{B} = \operatorname{rot}\boldsymbol{A}, \ \boldsymbol{E} = -\operatorname{grad}\varphi - \frac{1}{c}\dot{\boldsymbol{A}}. \tag{6.25}$$

The tensors of an electromagnetic field *in vacuo*.

$$F_{ik}=\left(\frac{\partial\Phi_k}{\partial x_i}-\frac{\partial\Phi_i}{\partial x_k}\right). \tag{6.28}$$

$$f_{ik}=F_{ik}=\begin{pmatrix} 0 & H_z & -H_y & -iE_x \\ H_z & 0 & H_x & -iE_y \\ H_y & -H_x & 0 & -iE_z \\ iE_x & iE_y & iE_x & 0 \end{pmatrix}. \tag{6.29}$$

The tensors of an electromagnetic field in matter:

$$f_{ik}=\begin{pmatrix} 0 & H_z & -H_y & -iD_x \\ H_z & 0 & H_x & -iD_y \\ H_y & H_x & 0 & -iD_z \\ iD_x & iD_y & iD_z & 0 \end{pmatrix},$$

$$F_{ik}=\begin{pmatrix} 0 & B_z & -B_y & -iE_x \\ -B_z & 0 & B_x & -iE_y \\ B_y & -B_x & 0 & -iE_z \\ iE_x & iE_y & iE_x & 0 \end{pmatrix}. \tag{6.31}$$

The tensor of the momenta:

$$\mathfrak{M}_{ik}=\begin{pmatrix} 0 & M_z & -M_y & iP_x \\ -M_z & 0 & M_y & iP_y \\ M_y & -M_x & 0 & iP_z \\ -iP_x & -iP_y & -iP_z & 0 \end{pmatrix}, \tag{6.33}$$

now $\boldsymbol{B}=\boldsymbol{H}+4\pi\boldsymbol{M}$ and $\boldsymbol{E}=\boldsymbol{D}-4\pi\boldsymbol{P}$.

The field transformation formulae:

$$\begin{aligned} E_x &= E_x, & B_x &= B'_x, \\ E_y &= \Gamma\left(E_y+\frac{V}{c}B'_z\right), & B_y &= \Gamma\left(B'_y-\frac{V}{c}E'_z\right), \\ E_z &= \Gamma\left(E'_z-\frac{V}{c}B'_y\right), & B_z &= \Gamma\left(B'_z+\frac{V}{c}E'_y\right); \end{aligned} \tag{6.36}$$

$$\begin{aligned} D_x &= D'_x, & H_x &= H'_x, \\ D_y &= \Gamma\left(D'_y+\frac{V}{c}H'_z\right), & H'_y &= \Gamma\left(H'_y-\frac{V}{c}D'_z\right), \\ D_z &= \Gamma\left(D'_z-\frac{V}{c}H'_y\right), & H_z &= \Gamma\left(H'_z+\frac{V}{c}D'_y\right). \end{aligned} \tag{6.37}$$

The same formulae expressed via the projections on the relative velocity direction and the direction perpendicular to it:

$$\boldsymbol{E}_{\parallel}=\boldsymbol{E}'_{\parallel},\quad \boldsymbol{E}_{\perp}=\Gamma\left(\boldsymbol{E}'_{\perp}-\frac{1}{c}[\boldsymbol{V}\boldsymbol{B}']\right),\quad \boldsymbol{B}_{\parallel}=\boldsymbol{B}'_{\parallel},$$
$$\boldsymbol{B}_{\perp}=\Gamma\left(\boldsymbol{B}'_{\perp}+\frac{1}{c}[\boldsymbol{V}\boldsymbol{E}']\right); \tag{6.38}$$

$$\boldsymbol{D}_{\parallel}=\boldsymbol{D}'_{\parallel},\quad \boldsymbol{D}_{\perp}=\Gamma\left(\boldsymbol{D}'_{\perp}-\frac{1}{c}[\boldsymbol{V}\boldsymbol{H}']\right),\quad \boldsymbol{H}_{\parallel}=\boldsymbol{H}'_{\parallel},$$
$$\boldsymbol{H}_{\perp}=\Gamma\left(\boldsymbol{H}'_{\perp}+\frac{1}{c}[\boldsymbol{V}\boldsymbol{D}']\right). \tag{6.42}$$

The Lorentz force density

$$\boldsymbol{f}=\rho\left\{\boldsymbol{E}+\frac{1}{c}[\boldsymbol{v}\boldsymbol{B}]\right\}. \tag{6.49}$$

The field invariants (§ 6.5):

$$I_1=F_{ik}^2=E^2-H^2,\quad I_2=2i\boldsymbol{E}\boldsymbol{H}.$$

The four-dimensional expression for the Lorentz force density is the same in SI and the Gaussian system:

$$f_i=\frac{1}{c}F_{ik}s_k. \tag{6.53}$$

The Maxwell equations in the three-dimensional form

$$\operatorname{rot}\boldsymbol{H}=\frac{4\pi}{c}\boldsymbol{j}+\frac{1}{c}\dot{\boldsymbol{D}},\quad \operatorname{div}\boldsymbol{D}=4\pi\rho, \tag{6.56}$$

$$\operatorname{rot}\boldsymbol{H}=-\frac{1}{c}\dot{\boldsymbol{B}},\qquad \operatorname{div}\boldsymbol{B}=0. \tag{6.57}$$

The Maxwell equations in the four-dimensional form:

$$\frac{\partial f_{ik}}{\partial x_k}=\frac{4\pi}{c}s_i, \tag{6.60}$$

$$\frac{\partial F_{ik}}{\partial x_l}+\frac{\partial F_{kl}}{\partial x_i}+\frac{\partial F_{li}}{\partial x_k}=0. \tag{6.67}$$

The material equations in the three-dimensional form do not vary:

$$\boldsymbol{D}'=\varepsilon\boldsymbol{E}', \tag{6.68}$$
$$\boldsymbol{B}'=\mu\boldsymbol{H}', \tag{6.69}$$
$$\boldsymbol{j}'=\sigma\boldsymbol{E}'. \tag{6.70}$$

The material relations in a moving medium:

$$\boldsymbol{D}+\frac{1}{c}[\boldsymbol{VH}]=\varepsilon\left(\boldsymbol{E}+\frac{1}{c}[\boldsymbol{VB}]\right), \tag{6.74}$$

$$\boldsymbol{B}-\frac{1}{c}[\boldsymbol{VE}]=\mu\left(\boldsymbol{H}-\frac{1}{c}[\boldsymbol{VD}]\right). \tag{6.75}$$

Solving these equations with respect to $\boldsymbol{D}$ and $\boldsymbol{B}$, we shall obtain ($\mathrm{B}=V/c$)

$$\boldsymbol{D}=\frac{1}{1-\varepsilon\mu\mathrm{B}^2}\left\{\varepsilon\boldsymbol{E}(1-\mathrm{B}^2)+(\varepsilon\mu-1)\left\{\frac{1}{c}[\boldsymbol{VH}]-\frac{\varepsilon}{c^2}\boldsymbol{V}(\boldsymbol{VE})\right\}\right.,$$

$$\boldsymbol{B}=\frac{1}{1-\varepsilon\mu\mathrm{B}^2}\left\{\mu\boldsymbol{H}(1-\mathrm{B}^2)-(\varepsilon\mu-1)\left\{\frac{1}{c}[\boldsymbol{VH}]-\frac{\mu}{c^2}\boldsymbol{V}(\boldsymbol{VE})\right\}\right.,$$

whence

$$\boldsymbol{D}_{\parallel}=\varepsilon\boldsymbol{E}_{\parallel},\quad \boldsymbol{B}_{\parallel}=\mu\boldsymbol{H}_{\parallel}, \tag{6.76}$$

$$\begin{aligned}(1-\varepsilon\mu\mathrm{B}^2)\boldsymbol{D}_{\perp}&=\varepsilon(1-\mathrm{B}^2)\boldsymbol{E}_{\perp}+(\varepsilon\mu-1)\frac{1}{c}[\boldsymbol{VH}],\\(1-\varepsilon\mu\mathrm{B}^2)\boldsymbol{B}_{\perp}&=\mu(1-\mathrm{B}^2)\boldsymbol{H}_{\perp}-(\varepsilon\mu-1)\frac{1}{c}[\boldsymbol{VE}].\end{aligned} \tag{6.77}$$

Ignoring the values of B^2 and $\varepsilon\mu\mathrm{B}^2$ as compared to unity in Eq. (6.77), we shall get

$$\begin{aligned}\boldsymbol{D}&=\varepsilon\boldsymbol{E}+\frac{1}{c}(\varepsilon\mu-1)[\boldsymbol{VH}],\\\boldsymbol{B}&=\mu\boldsymbol{H}-\frac{1}{c}(\varepsilon\mu-1)[\boldsymbol{VE}].\end{aligned} \tag{6.78}$$

The material equations in the four-dimensional form:

$$f_{ik}u_k=\varepsilon F_{ik}u_k, \tag{6.79}$$

$$F_{ik}u_l+F_{kl}u_i+F_{li}u_k=\mu(f_{ik}u_l+f_{kl}u_i+f_{li}u_k), \tag{6.80}$$

$$s_i=\frac{\sigma}{c}F_{ik}u_k. \tag{6.81}$$

The electromagnetic field energy density:

$$w=\frac{\boldsymbol{ED}+\boldsymbol{BH}}{8\pi}.$$

The Poynting-Umov vector:

$$\boldsymbol{S}=\frac{c}{4\pi}[\boldsymbol{EH}].$$

The energy-momentum-tension tensor *in vacuo*.

$$T_{ik}=\begin{pmatrix} T_{\alpha\beta} & -icg \\ -\frac{i}{c}S & w \end{pmatrix}, \quad g=\frac{S}{c^2}=\frac{1}{4\pi c}[EH].$$

$$T_{\alpha\beta}=\frac{1}{4\pi}\{E_\alpha D_\beta+H_\alpha B_\beta\}-\delta_{\alpha\beta}(w).$$

The energy-momentum-tension tensors in matter (Minkowski and Abraham):

$$T_{ik}^{M}=\begin{pmatrix} T_{\alpha\beta}^{M} & -icg^{M} \\ -\frac{i}{c}S & w \end{pmatrix}, \quad g^{M}=\frac{1}{4\pi c}[DB],$$

$$T_{\alpha\beta}^{M}=\frac{1}{4\pi}\{E_\alpha D_\beta+H_\alpha B_\beta\}-\delta_{\alpha\beta}w,$$

$$T_{ik}^{A}=\begin{pmatrix} T_{\alpha\beta}^{A} & -icg^{A} \\ -\frac{i}{c}S & w \end{pmatrix}, \quad g^{A}=\frac{S}{c^2}=\frac{1}{4\pi c}[EH],$$

$$T_{\alpha\beta}^{A}=\frac{1}{8\pi}\{E_\alpha D_\beta+E_\beta D_\alpha+H_\alpha B_\beta+H_\beta B_\alpha\}-\delta_{\alpha\beta}w.$$

BIBLIOGRAPHY

This list includes only those books which cover the subject of the theory of relativity very substantially. Brief comments given for the readers' convenience reflect the personal views of the author of this book.

1. *The Principle of Relativity*, a collection of articles by the classics of the relativistic theory, Moscow-Leningrad, 1935 (in Russian).

The original articles by Lorentz, Poincaré, Einstein and Minkowski. The articles by the first two authors make it possible to form an opinion about the immediate forerunners of Einstein. The article by Einstein "On the Electrodynamics of Moving Bodies" covers the special theory of relativity almost completely, except for the problems of thermodynamics Minkowski's report presents the fundamentals of the STR in terms of four-dimensional geometrical physics.

2. *The Principle of Relativity*, a collection of articles on the special theory of relativity, Moscow, 1973 (in Russian).

This book contains all the material on the STR published in the previous collection of 1935. Some more articles by Poincaré, Lorentz, Planck and Pauli are added. The third part of the book is devoted to the history of the STR.

3. A. Einstein, *Collected Works*, Vol. I, Moscow, 1965; Vol. II, 1966; Vol. IV, 1967 (in Russian).

Volumes I and II include Einstein's works on the theory of relativity. Volume I contains papers on both the STR and the general theory of relativity. Volume II covers basically the articles on the general theory of relativity. Volume I contains the translations of two works by Einstein published earlier as separate editions (see refs. 4 and 5 below).

4. A Einstein, *The Meaning of Relativity*, Princeton, 1953.

A very concise and rather complicated presentation of the fundamental ideas.

5. A. Einstein. *Über die spezielle und die allgemeine Relativitätstheorie (gemeinverständlich)*, Braunschweig, 1920.

Once Einstein joked that this "comprehensible exposition" (gemeinverständlich) should be called "incomprehensible" exposition.

6. L I Mandelshtam, Lectures on Physical Foundations of the Theory of Relativity. *Complete Works*, Vol. V, Moscow, 1950 A separate publication· Lectures on Optics, Theory of Relativity and Quantum Mechanics, Moscow, 1972 (in Russian).

The lectures given by L. I. Mandelshtam and prepared for publication by his disciples deal with the history of the special theory of relativity and touch upon the major problems of the theory. The lectures were intended for an advanced audience and cannot be recommended to the reader who is new to the subject. The sophisticated reader, however, should get acquainted with these lectures by all means.

7. C. Moeller, *The Theory of Relativity*, Oxford, 1952.

An extended course of university lectures covers a wide range of problems of the STR and the GTR. It requires deep knowledge of physics and mathematics.

8. W. Pauli, *Theory of Relativity*, London, 1958.
The book is the translation of an article from the Mathematical Encyclopedia written in 1921. The presentation of the subject is fairly complete, even though concise. The article has a great number of references.

9. L. Landau, E. Lifshits, *The Field Theory*, the last 6th ed. Moscow, 1973 (in Russian).
The first four chapters of the book are in fact devoted to the STR. By and large, the book systematically presents the general and special theory of relativity in terms of the least action principle. Fields in media are not considered in this book; they are treated by the same authors in the book *Continuum Electrodynamics*, Moscow, 1967.
The book is intended for readers who have a good grounding, for the exposition of some of the themes is rather concise; many details needed for the first acquaintance with the STR are omitted. One of the best books on the STR as concerns both factual data and the manner of presentation.

10 R. C. Tolman, *Relativity, Thermodynamics and Cosmology*, Oxford, 1934.
A classical work devoted both to the STR and to the GTR. It is written thoroughly and deliberately and is distinguished for the wide scope of the problems treated.

11. E. F. Taylor, J. A. Wheeler, *Spacetime Physics*, San Francisco-London, 1966.
The book is remarkable for an attempt made by the authors to present the STR at the very beginning of a college course. It presents the relativistic concepts of space and time as well as the relativistic mechanics in much detail. The book is partially based on the lectures given by Wheeler at the refresher courses for US teachers. It contains many superb illustrations, drawings and diagrams; some chapters are provided with supplements including about a hundred very useful problems pertaining to the most delicate themes of the STR. The book is concluded with a brief exposition of the general theory of relativity.

12. P. G. Bergmann, *Introduction to the Theory of Relativity*, N.Y., 1946.
The book describes first the STR and then the GTR occupying the larger part. The first nine chapters are devoted to the exposition of the STR. While trying at the very beginning to prepare the mathematical background for the GTR, the author overcomplicated the presentation of the STR.

13. R. Becker, *Theorie der Electrizität*, Stuttgart, 1957.
A very clear and popular presentation of the principal problems of the STR; the book is quite understandable for a student versed in electrodynamics

14. W. Panofsky, M. Phillips, *Classical Electricity and Magnetism*, New York, 1955
In the second half of the book the reader will find an up-to-date and interesting presentation of the STR; every chapter is supplemented with problems and references

15. A Sommerfeld, *Lectures on Theoretical Physics*, Vol. 3, Electrodynamics, New York, 1952.
The third chapter of the book is devoted to the detailed study of the theory of relativity and the electron theory. A rather peculiar and sufficiently sophisticated exposition, of great interest to a prepared reader

16. V. A. Fok, *Theory of Space, Time and Gravitation*, Moscow, 1964 (in Russian).
Although the book is primarily concerned with the GTR, the first chapters are devoted to the STR. A number of problems are stated in an original manner.

17. J. D. Jackson, *Classical Electrodynamics*, New York-London, 1962.
A modern course of electrodynamics. Chapter 2 is devoted to the brief exposition of the STR. The applications are considered in Ch. 12 and partially in Ch. 14. These chapters are supplied with problems.

18. M. A Tonnelat, *Les principes de la théorie électromagnétiques et de la relativité*, Paris, 1959.
The book gives a detailed exposition of many questions of the STR and the GTR; however, the presentation is overly complicated in notation and mathematics.

19. *The Feynman Lectures on Physics*, R. P. Feynman, R. B Leighton. M Sands, Cambridge (Mass), 1967.
The lectures delivered by Feynman is an interesting endevour to combine the general course of physics with the course of theoretical physics. The theory of relativity is dwelt upon in some chapters of this collection.

20. D. Bohm, *The Special Theory of Relativity*, New York-Amsterdam, 1965
A very interesting book stressing the basic philosophical problems of the STR. The principal subject of the book is the development of physical concepts of space and time The concrete problems of the STR touched upon in the book are presented in a clearcut and precise way. The book requires critical attitude when the philosophical problems are discussed. It is intended for a sophisticated reader.

21. H. Bondi, *Assumption and Myth in Physical Theory*, Cambridge, 1967
This small booklet discusses the fundamental problems of the STR and the gravitational theory. It is very useful for the first acquaintance with the STR.

22. H. Bondi, *Relativity and Common Sense*, London, 1965.
The book is remarkable for its methodical approach to the presentation of the STR. It justly points out that the presentation of the STR from the standpoint of the 19th century physics is not expedient now In this connection the Michelson experiment is evaluated from the position of contemporary physics. At the same time, a new method of presentation of the STR is suggested which introduces the observers possessing identical clocks and radars. Thus, the use of rigid scales is obviated. The *K* calculus allows the basic kinematic results of the STR to be obtained directly from the Einstein postulates, bypassing the Lorentz transformation. Such a presentation extensively uses geometrical diagrams, the author shows how the *K* calculus can be linked with the Lorentz transformation as well as with the conventional ways of presentation.

23. P. G. Bergmann, *The Riddle of Gravitation*, New York, 1968.
This small book containing no technical details clearly points out the limitedness of the special theory of relativity and the inevitable transition in the general case to the GTR. It indicates those conclusions of the STR that served as a starting point for the construction of the gravitational theory. *The Riddle of Gravitation* consists of three parts: one devoted to the STR, the second to the GTR, and the third, containing the most recent data of astronomical observations The reader of the book is supposed to be acquainted with the technical college courses in physics and mathematics. The first part of the book cannot, however, be used for the first acquaintance with the STR, but is an excellent summary to be utilized for recapitulation

24. C. Lanczos, *Albert Einstein and the Cosmic World Order*, New York, 1965.
This book can by no means serve as a textbook on the STR However, the principal approach to the construction of the STR is described clearly and precisely. The first three chapters of the book should be recommended to everyone who begins to study the STR.

25. Yu. B. Rumer, M. S. Ryvkin, *The Theory of Relativity*, Moscow, 1960 (in Russian).
A comparatively simple presentation of the principal ideas and results of the STR; the book is intended for physics departments of pedagogical colleges.

26. B. Jaffe, *Michelson and the Speed of Light*, London, 1961.

27. V. A Ugarov, Photographing Bodies Moving at Relativistic Velocities, *The Einstein Collection of 1973*, Moscow, 1974 (in Russian).

28. Ya. A. Smorodinsky, V. A. Ugarov, Two Paradoxes of the Special Theory of Relativity, UFN, 107, 141 (1972).

29. I. E. Tamm, *Fundamentals of the Theory of Electricity*, Mir Publishers, Moscow, 1979

30 J. A. Stratton, *Electromagnetic Theory*, New York-London, 1941,

31. L. Marder, *Time and Space-Traveller*, London, 1971.

32. Ya. P. Terletsky, *The Paradoxes of the Theory of Relativity*, Moscow, 1966 (in Russian).

33. V. L. Ginzburg, *Theoretical Physics and Astrophysics*, Moscow, 1975 (in Russian).

34 H. Goldstein, *Classical Mechanics*, Cambridge (Mass), 1950.

35 C. Lanczos, *The Variational Principles of Mechanics*, Toronto, 1966

36. A. Sommerfeld, *Lectures on Theoretical Physics*, Vol. 4, Optics, New York, 1954.

INDEX

TO THE READER

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make.

Our address is:
USSR, 129820, I-110, GSP
Pervy Rizhsky Pereulok 2
Mir Publishers

Printed in the Union of Soviet Socialist Republics

THE THEORY OF PROBABILITY

by B GNEDENKO

The book presents the fundamentals of probability theory, the mathematical science that deals with the laws of random phenomena. These laws play an extremely important role in physical and other fields of natural science, in engineering, economics, linguistics and so forth.

The material covered ranges over the following problems: the concept of probability, sequences of independent trials, Markov chains, random variables and distribution functions, numerical characteristics of random variables, the law of large numbers, characteristic functions, the classical limit theorems, the theory of infinitely divisible distribution laws, the theory of stochastic processes, and elements of the theory of queues

The theory of probability is presented as a mathematical discipline, however the examples given not only illustrate the general propositions of the theory but provide links with problems that occur in the natural sciences.

The theory of probability is a text for students of mathematical departments of colleges and universities. It will also be found of definite interest to specialists in a wide range of fields (physicists, engineers, economists, linguists and others) that the science of probability touches on

MIR PUBLISHERS MOSCOW